GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:38:23 ; Search time 38 Seconds
(without alignments)
995.873 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1525	100.0	284	22	AAB62695	ABC1 protein exter
2	1525	100.0	2143	21	AAB38108	Human ABC1 cholest
3	1525	100.0	2259	21	AAB38107	Human ABC1 FHA-3 m
4	1525	100.0	2260	21	AAB38106	Human ABC1 cholest
5	1525	100.0	2261	21	AAB38082	Human ABC1 cholest
6	1525	100.0	2261	21	AAB38105	Human ABC1 cholest
7	1525	100.0	2261	21	AAB38109	Human ABC1 cholest
8	1525	100.0	2261	21	AAB38110	Human ABC1 cholest
9	1525	100.0	2261	21	AAB38111	Human ABC1 cholest
10	1525	100.0	2261	21	AAB38112	Human ABC1 cholest
11	1525	100.0	2261	21	AAB38113	Human ABC1 cholest
12	1525	100.0	2261	21	AAB38114	Human ABC1 cholest
13	1525	100.0	2261	21	AAB38116	Human ABC1 cholest
14	1525	100.0	2261	21	AAB38117	Human ABC1 cholest
15	1525	100.0	2261	22	AAB71749	Human ABC1 protein
16	1525	100.0	2261	22	AAB31361	Amino acid sequenc
17	1525	100.0	2261	22	AAB31362	Amino acid sequenc
18	1525	100.0	2261	22	AAB31363	Amino acid sequenc
19	1525	100.0	2261	22	AAB31365	Amino acid sequenc
20	1525	100.0	2261	22	AAB31366	Amino acid sequenc
21	1525	100.0	2261	22	AAB31367	Amino acid sequenc
22	1522	99.8	2261	21	AAB38115	Human ABC1 cholest
23	1513	99.2	2130	22	AAU02190	Human ABC1 mutant
24	1513	99.2	2201	21	AAY79380	Human ATP binding
25	1513	99.2	2201	22	AAE13021	Human ATP binding
26	1513	99.2	2201	22	AAM50227	Human ATP binding
27	1513	99.2	2261	21	AAB38104	Human ABC1 cholest
28	1513	99.2	2261	22	AAE13022	Human ATP binding
29	1513	99.2	2261	22	AAM50228	Human ATP binding
30	1513	99.2	2261	22	AAM78550	Human protein SEQ
31	1513	99.2	2261	22	AAU02176	Human ABC1. Homo
32	1513	99.2	2261	22	AAU02177	Human ABC1 mutant
33	1513	99.2	2261	22	AAU02181	Human ABC1 mutant
34	1513	99.2	2261	22	AAU02182	Human ABC1 mutant
35	1513	99.2	2261	22	AAU02183	Human ABC1 mutant
36	1513	99.2	2261	22	AAU02186	Human ABC1 mutant
37	1513	99.2	2261	22	AAU02189	Human ABC1 mutant
38	1513	99.2	2261	23	ABB81578	Human ABC-A-1-1 pr
39	1513	99.2	2261	23	AAE23000	Human ABC1 full-le
40	1513	99.2	2263	22	ABB11956	Human ABCA1 homolo
41	1513	99.2	2263	22	AAM79534	Human protein SEQ
42	1510	99.0	2261	22	AAU02188	Human ABC1 mutant
43	850	55.7	1525	22	AAU02187	Human ABC1 mutant
44	733.5	48.1	2273	19	AAW70398	ATP binding casset
45	663.5	43.5	1873	22	AAU04484	Human PD-ATP-bindi

ALIGNMENTS

RESULT 1
AAB62695
ID AAB62695 standard; peptide; 284 AA.
XX
AC AAB62695;
XX
DT 06-AUG-2001 (first entry)
XX
DE ABC1 protein external domain TM7-TM8 fragment (residues 1371-1654).
XX
KW ABC1; antilipemic; cholesterol; inhibitor; low density lipoprotein; LDL.
XX
OS Homo sapiens.
XX
PN WO200132184-A2.
XX
PD 10-MAY-2001.
XX
PF 01-NOV-2000; 2000WO-US30109.
XX
PR 01-NOV-1999; 99US-0162803.
PR 30-JUN-2000; 2000US-0215564.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Attie AD, Cook M, Gray-Keller MP, Hayden MR, Pimstone S;
PI Brooks-Wilson A;
XX
DR WPI; 2001-335779/35.
XX
PT New method for inhibiting cholesterol uptake in the gut comprises
PT administration of an inhibitor of an ABC1 protein -
XX

PS Disclosure; Page 9; 41pp; English.
XX
CC The invention relates to a new method for inhibiting cholesterol uptake
CC in the gut that comprises administration of an inhibitor of an ABC1
CC protein. The method is useful for: lowering levels of LDL (low density
CC lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
CC intestinal cells and the abundance of the ABC1 protein in the individual
CC by inhibiting the activity of the protein; identifying drugs that can
CC lower serum cholesterol and LDL levels comprises assaying the drug to
CC test if it can bind to an ABC1 protein; testing LDL cholesterol lowering
CC agents; and for modulation of ABC1 biological activity. Sequences
CC AAB62692-97 represent predicted external domain of ABC1 protein.
XX
SQ Sequence 284 AA;

Query Match 100.0%; Score 1525; DB 22; Length 284;
Best Local Similarity 100.0%; Pred. No. 2.3e-144;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
```

RESULT 2
AAB38108
ID AAB38108 standard; Protein; 2143 AA.
XX
AC AAB38108;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter FHA-1 mutant protein (R2144STOP).
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
DR N-PSDB; AAC69389.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2143 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2143;
Best Local Similarity 100.0%; Pred. No. 4.7e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 3
AAB38107
ID   AAB38107 standard; Protein; 2259 AA.
XX
AC   AAB38107;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 FHA-3 mutant protein (delta-E1893, D1894).
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-IB00532.
XX
PR   15-MAR-1999;   99US-0124702.
PR   08-JUN-1999;   99US-0138048.
PR   17-JUN-1999;   99US-0139600.
PR   01-SEP-1999;   99US-0151977.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON BIORESEARCH INC.
XX
PI   Hayden MR,  Wilson AR,  Pimstone SN;
XX
DR   WPI; 2000-587528/55.
DR   N-PSDB; AAC69388.
XX
PT   New ABC1 polypeptide is useful for treating diseases associated with
PT   ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2259 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2259;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 4
AAB38106
ID   AAB38106 standard; Protein; 2260 AA.
XX
AC   AAB38106;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter FHA-1 mutant protein (delta-L693).
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
```

PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
DR N-PSDB; AAC69387.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2260 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2260;
Best Local Similarity 100.0%; Pred. No. 5e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db 1370 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1429

Qy 61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
Db 1430 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1489

Qy 121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
Db 1490 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1549

Qy 181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
Db 1550 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1609

Qy 241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
Db 1610 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1653

RESULT 5
AAB38082
ID AAB38082 standard; Protein; 2261 AA.
XX
AC AAB38082;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
DR N-PSDB; AAC69120.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Claim 5; Page 152-157; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular

CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents the human ABC1 cholesterol transporter.
XX
SQ Sequence 2261 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db 1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy 61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
Db 1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy 121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
Db 1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy 181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
Db 1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy 241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
Db 1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654

RESULT 6
AAB38105
ID AAB38105 standard; Protein; 2261 AA.
XX
AC AAB38105;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter TD-2 mutant protein (Q597R).
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
DR N-PSDB; AAC69386.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2261 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db 1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy 61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 7
AAB38109
ID   AAB38109 standard; Protein; 2261 AA.
XX
AC   AAB38109;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter mutant, R219K.
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-IB00532.
XX
PR   15-MAR-1999;   99US-0124702.
PR   08-JUN-1999;   99US-0138048.
PR   17-JUN-1999;   99US-0139600.
PR   01-SEP-1999;   99US-0151977.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON BIORESEARCH INC.
XX
PI   Hayden MR,  Wilson AR,  Pimstone SN;
XX
DR   WPI; 2000-587528/55.
XX
PT   New ABC1 polypeptide is useful for treating diseases associated with
PT   ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 8
AAB38110
ID   AAB38110 standard; Protein; 2261 AA.
XX
AC   AAB38110;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter mutant, V399A.
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
```

KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2261 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654
```

RESULT 9
AAB38111
ID AAB38111 standard; Protein; 2261 AA.
XX
AC AAB38111;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter mutant, V771M.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is

```
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 10
AAB38112
ID   AAB38112 standard; Protein; 2261 AA.
XX
AC   AAB38112;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter mutant, T774P.
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-IB00532.
XX
PR   15-MAR-1999;   99US-0124702.
PR   08-JUN-1999;   99US-0138048.
PR   17-JUN-1999;   99US-0139600.
PR   01-SEP-1999;   99US-0151977.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON BIORESEARCH INC.
XX
PI   Hayden MR,  Wilson AR,  Pimstone SN;
XX
DR   WPI; 2000-587528/55.
XX
PT   New ABC1 polypeptide is useful for treating diseases associated with
PT   ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;
```

```
  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 11
AAB38113
ID   AAB38113 standard; Protein; 2261 AA.
XX
AC   AAB38113;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter mutant, K776N.
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-IB00532.
XX
PR   15-MAR-1999;   99US-0124702.
PR   08-JUN-1999;   99US-0138048.
PR   17-JUN-1999;   99US-0139600.
PR   01-SEP-1999;   99US-0151977.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON BIORESEARCH INC.
XX
PI   Hayden MR,  Wilson AR,  Pimstone SN;
XX
DR   WPI; 2000-587528/55.
XX
PT   New ABC1 polypeptide is useful for treating diseases associated with
PT   ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 12
AAB38114
ID   AAB38114 standard; Protein; 2261 AA.
XX
AC   AAB38114;
XX
DT   29-JAN-2001  (first entry)
XX
```

DE Human ABC1 cholesterol transporter mutant, E1172D.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples; Page -; 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2261 AA;

Query Match 100.0%; Score 1525; DB 21; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-143;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654
```

RESULT 13
AAB38116
ID AAB38116 standard; Protein; 2261 AA.
XX
AC AAB38116;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter mutant, S1731C.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN;
XX
DR WPI; 2000-587528/55.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's

```
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 14
AAB38117
ID   AAB38117 standard; Protein; 2261 AA.
XX
AC   AAB38117;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human ABC1 cholesterol transporter mutant, I883M.
XX
KW   Human ABC1 cholesterol transporter; chromosome 9q31;
KW   ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW   Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW   cardiovascular disease; coronary artery disease; coronary restenosis;
KW   cerebrovascular disease; peripheral vascular disease;
KW   Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW   X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW   prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW   mutein.
XX
OS   Homo sapiens.
XX
PN   WO200055318-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-IB00532.
XX
PR   15-MAR-1999;   99US-0124702.
PR   08-JUN-1999;   99US-0138048.
PR   17-JUN-1999;   99US-0139600.
PR   01-SEP-1999;   99US-0151977.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON BIORESEARCH INC.
XX
PI   Hayden MR,  Wilson AR,  Pimstone SN;
XX
DR   WPI; 2000-587528/55.
XX
PT   New ABC1 polypeptide is useful for treating diseases associated with
PT   ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT   disease and cancer -
XX
PS   Examples; Page -; 229pp; English.
XX
CC   The invention relates to the human ABC1 cholesterol transporter protein
CC   (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC   a member of the ATP-binding cassette (ABC transporter) superfamily of
CC   proteins, and plays a crucial role in cholesterol transport, particularly
CC   intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC   involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC   located on chromosome 9q31, and mutations in this gene are associated
CC   with two genetic HDL (high density lipoprotein) deficiency disorders,
CC   Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC   are distinguishable in that TD is an autosomal recessive disorder, while
CC   FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC   cholesterol") in the blood correlate with a high risk of cardiovascular
CC   disease, particularly coronary artery disease, but also cerebrovascular
CC   disease, coronary restenosis, and peripheral vascular disease.
CC   Conversely, a high level of HDL has protective effects against
CC   cardiovascular disease. The invention provides genetic constructs and
CC   transgenic cells and non-human animals comprising human ABC1 nucleic
CC   acids, and methods of gene therapy for the treatment or prevention of
CC   cardiovascular disease comprising the administration of an expression
CC   vector encoding ABC1 or an active fragment thereof. The invention also
CC   encompasses compounds which mimic ABC1 activity, compounds which
CC   stimulate ABC1 expression and methods of screening for such compounds.
CC   It further relates to methods for determining whether a patient has an
CC   increased risk for cardiovascular disease due to polymorphisms in the
CC   ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC   or prevent cardiovascular disease, especially coronary artery disease,
CC   cerebrovascular disease, coronary restenosis or peripheral vascular
CC   disease. They may also be used in the treatment of diseases associated
CC   with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC   disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
```

```
CC   The invention specifically excludes proteins with the exact amino acid
CC   sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC   acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC   present sequence represents a mutant human ABC1 cholesterol transporter
CC   associated with an altered cholesterol level and therefore an altered
CC   risk of cardiovascular disease.
CC   Note: The present sequence is not shown in the specification, but is
CC   derived from the native human ABC1 shown on pages 152-157.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 21;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 15
AAB71749
ID   AAB71749 standard; protein; 2261 AA.
XX
AC   AAB71749;
XX
DT   17-MAY-2001  (first entry)
XX
DE   Human ABC1 protein.
XX
KW   High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1.
XX
OS   Homo sapiens.
XX
PN   WO200115676-A2.
XX
PD   08-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-IB01492.
XX
PR   01-SEP-1999;   99US-0151977.
PR   15-MAR-2000; 2000US-0526193.
PR   23-JUN-2000; 2000US-0213958.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (XENO-) XENON GENETICS INC.
XX
PI   Hayden MR,  Brooks-Wilson AR,  Pimstone SN,  Clee SM;
XX
DR   WPI; 2001-244356/25.
XX
PT   Treating a lower than normal high density lipoprotein-cholesterol
PT   (HDL-C) level, a higher than normal triglyceride level, or a
PT   cardiovascular disease, by administering a compound that modulates LXR-
PT   or RXR-mediated transcriptional activity -
XX
PS   Claim 16; Fig 2; 317pp; English.
XX
CC   The present invention relates to a method for treating a patient
CC   diagnosed as having a lower than normal high density
CC   lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC   triglyceride level, or a cardiovascular disease, involving
CC   administering a compound that modulates LXR- or RXR-mediated
CC   transcriptional activity or ABC1 expression or activity.
CC   The LXR gene product may be used in an assay to identify
CC   compounds useful for the treatment of a disease or condition selected a
CC   lower than normal HDL cholesterol level, a higher than normal
CC   triglyceride level, and a cardiovascular disease.
XX
SQ   Sequence   2261 AA;

  Query Match             100.0%;  Score 1525;  DB 22;  Length 2261;
  Best Local Similarity   100.0%;  Pred. No. 5e-143;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


Search completed: February  4, 2003, 09:39:19
Job time : 48 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:39:23 ; Search time 14 Seconds
(without alignments)
596.865 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1423	93.3	1375	3	US-08-665-259-26	Sequence 26, Appl
2	1423	93.3	1375	3	US-08-762-500-26	Sequence 26, Appl
3	257	16.9	1457	3	US-08-665-259-27	Sequence 27, Appl
4	257	16.9	1457	3	US-08-762-500-27	Sequence 27, Appl
5	94.5	6.2	884	6	5208144-8	Patent No. 5208144
6	92	6.0	2511	4	US-09-261-907-2	Sequence 2, Appli
7	89.5	5.9	596	4	US-08-481-190-8	Sequence 8, Appli
8	89.5	5.9	596	5	PCT-US93-00869-8	Sequence 8, Appli
9	89	5.8	903	3	US-08-804-439A-22	Sequence 22, Appl
10	89	5.8	903	3	US-08-720-229-22	Sequence 22, Appl
11	88.5	5.8	903	1	US-08-220-151-8	Sequence 8, Appli
12	88.5	5.8	903	1	US-08-413-118-8	Sequence 8, Appli
13	88.5	5.8	903	3	US-08-473-446-8	Sequence 8, Appli
14	88	5.8	888	2	US-08-861-464-6	Sequence 6, Appli
15	88	5.8	888	2	US-08-396-001-6	Sequence 6, Appli
16	88	5.8	888	4	US-09-323-433A-6	Sequence 6, Appli
17	86.5	5.7	967	4	US-09-139-802-201	Sequence 201, App
18	86	5.6	2509	1	US-08-469-005A-10	Sequence 10, Appl
19	85.5	5.6	953	4	US-09-245-281-43	Sequence 43, Appl
20	85.5	5.6	953	4	US-09-207-359B-43	Sequence 43, Appl
21	83.5	5.5	339	4	US-09-125-619-14	Sequence 14, Appl
22	83.5	5.5	969	4	US-09-206-942-32	Sequence 32, Appl
23	83.5	5.5	975	4	US-09-206-942-30	Sequence 30, Appl
24	83.5	5.5	3224	2	US-08-705-660-34	Sequence 34, Appl
25	83.5	5.5	3224	3	US-08-989-045-34	Sequence 34, Appl
26	83	5.4	913	1	US-08-220-151-6	Sequence 6, Appli
27	83	5.4	913	1	US-08-413-118-6	Sequence 6, Appli
28	83	5.4	913	3	US-08-473-446-6	Sequence 6, Appli
29	83	5.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
30	82.5	5.4	736	1	US-07-688-352C-24	Sequence 24, Appl
31	82.5	5.4	736	2	US-08-474-379C-24	Sequence 24, Appl
32	82.5	5.4	736	3	US-09-146-249A-24	Sequence 24, Appl
33	82.5	5.4	736	3	US-08-206-188B-24	Sequence 24, Appl
34	82.5	5.4	736	5	PCT-US91-02714-23	Sequence 23, Appl
35	82	5.4	904	6	5244792-4	Patent No. 5244792
36	82	5.4	913	6	5196516-8	Patent No. 5196516
37	81.5	5.3	820	4	US-09-173-914-2	Sequence 2, Appli
38	81.5	5.3	885	1	US-08-042-747A-8	Sequence 8, Appli
39	81.5	5.3	885	3	US-08-804-439A-23	Sequence 23, Appl
40	81.5	5.3	885	3	US-08-720-229-23	Sequence 23, Appl
41	80	5.2	464	4	US-09-025-580-28	Sequence 28, Appl
42	80	5.2	464	4	US-09-457-040B-5	Sequence 5, Appli
43	80	5.2	605	2	US-08-752-307B-8	Sequence 8, Appli
44	80	5.2	605	4	US-09-707-802-8	Sequence 8, Appli
45	80	5.2	605	4	US-09-991-326-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-08-665-259-26
; Sequence 26, Application US/08665259
; Patent No. 6028173
;  GENERAL INFORMATION:
;    APPLICANT:  Landes, Gregory M.
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Connors, Timothy D.
;    APPLICANT:  Dackowski, William R.
;    APPLICANT:  Van Raay, Terence J.
;    APPLICANT:  Klinger, Katherine W.
;    TITLE OF INVENTION:  NOVEL HUMAN CHROMOSOME 16 GENES,
;    TITLE OF INVENTION:  COMPOSITIONS, METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  One Mountain Road
;      CITY:  Framingham
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,259
;      FILING DATE:  17-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dugan, Deborah A.
;      REGISTRATION NUMBER:  37,315
;      REFERENCE/DOCKET NUMBER:  IG5-9.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 872-8400
;      TELEFAX:  (508) 872-5415
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1375 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-665-259-26

  Query Match           93.3%;  Score 1423;  DB 3;  Length 1375;
  Best Local Similarity  93.0%;  Pred. No. 6.2e-141;
  Matches  264;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;
```

```
Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db    485 FGKYPSLELQPWMYNEQYTFVSNDAPEDMGTQELLNALTKDPGFGTRCMEGNPIPDTPCL 544

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||:|| :|||||:|:||||||||||:|||||||||||||||||||||||||||||||
Db    545 AGEEDWTISPVPQSIVDLFQNGNWTMKNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 604

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | ||||||:|||||||||||||||||||||||||||||||||||||||:|||||| |
Db    605 QKTADILQNLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHE 664

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          |||||||||| ||| ||:||||||:|||||| ||||:|||||||||||||||||||||||
Db    665 VNDAIKQMKKLLKLTKDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLNVIN 724

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||| ||||||||||||||||||||||||||||
Db    725 NAILRANLQKGENPSQYGITAFNHPLNLTKQQLSEVALMTTSVD 768


RESULT 2
US-08-762-500-26
; Sequence 26, Application US/08762500
; Patent No. 6030806
;  GENERAL INFORMATION:
;    APPLICANT:  Landes, Gregory M.
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Connors, Timothy D.
;    APPLICANT:  Dackowski, William R.
;    APPLICANT:  Van Raay, Terence J.
;    APPLICANT:  Klinger, Katherine W.
;    TITLE OF INVENTION:  NOVEL HUMAN CHROMOSOME 16 GENES,
;    TITLE OF INVENTION:  COMPOSITIONS, METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  83
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  One Mountain Road
;      CITY:  Framingham
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/762,500
;      FILING DATE:  09-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/665,259
;      FILING DATE:  17-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10469
;      FILING DATE:  17-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dugan, Deborah A.
;      REGISTRATION NUMBER:  37,315
;      REFERENCE/DOCKET NUMBER:  IG5-9.3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 872-8400
;      TELEFAX:  (508) 872-5415
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1375 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-762-500-26

  Query Match             93.3%;  Score 1423;  DB 3;  Length 1375;
  Best Local Similarity   93.0%;  Pred. No. 6.2e-141;
  Matches  264;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db    485 FGKYPSLELQPWMYNEQYTFVSNDAPEDMGTQELLNALTKDPGFGTRCMEGNPIPDTPCL 544

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||:|| :|||||:|:||||||||||:|||||||||||||||||||||||||||||||
Db    545 AGEEDWTISPVPQSIVDLFQNGNWTMKNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 604

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | ||||||:|||||||||||||||||||||||||||||||||||||||:|||||| |
Db    605 QKTADILQNLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHE 664

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          |||||||||| ||| ||:||||||:|||||| ||||:|||||||||||||||||||||||
Db    665 VNDAIKQMKKLLKLTKDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLNVIN 724

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||||||| ||||||||||||||||||||||||||||
Db    725 NAILRANLQKGENPSQYGITAFNHPLNLTKQQLSEVALMTTSVD 768


RESULT 3
US-08-665-259-27
; Sequence 27, Application US/08665259
; Patent No. 6028173
;  GENERAL INFORMATION:
;    APPLICANT:  Landes, Gregory M.
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Connors, Timothy D.
;    APPLICANT:  Dackowski, William R.
;    APPLICANT:  Van Raay, Terence J.
;    APPLICANT:  Klinger, Katherine W.
;    TITLE OF INVENTION:  NOVEL HUMAN CHROMOSOME 16 GENES,
;    TITLE OF INVENTION:  COMPOSITIONS, METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  One Mountain Road
;      CITY:  Framingham
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,259
;      FILING DATE:  17-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dugan, Deborah A.
;      REGISTRATION NUMBER:  37,315
;      REFERENCE/DOCKET NUMBER:  IG5-9.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 872-8400
;      TELEFAX:  (508) 872-5415
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-665-259-27

  Query Match             16.9%;  Score 257;  DB 3;  Length 1457;
  Best Local Similarity   24.9%;  Pred. No. 8.6e-18;
```

```
  Matches   91;  Conservative   40;  Mismatches   98;  Indels  136;  Gaps   15;

Qy       2 GKYPSLELQPWMYNEQYT--------FVSNDAPE-------DTGTLELLNALTKDPGFGT 46
           |  | | | |  |:  ||        : : :  |       |    :|::      | |
Db     504 GDLPPLVLSPSQYH-NYTQPRGNFIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGA 562

Qy      47 RCM-----------------------------------EGNPI--------------- 54
            |:                                   :| |:
Db     563 TCVLKSPANGSLGPMLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDS 622

Qy      55 ---PD----------TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKI 100
              ||           |  || | ||:|| :|: : :  :            | ||:
Db     623 PVXPDEDSLQAWNMSLPPTAGPETWTSAPSLPRLVHEPVR-----------CTCSAQGT 670

Qy     101 KKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEF 160
               ||   ||  ||| :  | ||| |:|| |:|:||: |  :
Db     671 GFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----------RLH 715

Qy     161 RYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
           |||  : |  | |   |:                    |   :   :  :  |
Db     716 RYGAITFG--NVQKSIPAS---------------------FGARVPPMVRKIAVRRVA 750

Qy     221 KVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQLS-EVAL 278
           :| :||||:|:: ::|| :|||||||||| | : ||: | ||  |||:| |   || :  |
Db     751 QVLYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYXITVTNHPMNKTSASLSLDYLL 810

Qy     279 MTTSV 283
             | |
Db     811 QGTDV 815


RESULT 4
US-08-762-500-27
; Sequence 27, Application US/08762500
; Patent No. 6030806
;  GENERAL INFORMATION:
;    APPLICANT:  Landes, Gregory M.
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Connors, Timothy D.
;    APPLICANT:  Dackowski, William R.
;    APPLICANT:  Van Raay, Terence J.
;    APPLICANT:  Klinger, Katherine W.
;    TITLE OF INVENTION:  NOVEL HUMAN CHROMOSOME 16 GENES,
;    TITLE OF INVENTION:  COMPOSITIONS, METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  83
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  One Mountain Road
;      CITY:  Framingham
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/762,500
;      FILING DATE:  09-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/665,259
;      FILING DATE:  17-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10469
;      FILING DATE:  17-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dugan, Deborah A.
;      REGISTRATION NUMBER:  37,315
;      REFERENCE/DOCKET NUMBER:  IG5-9.3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 872-8400
;      TELEFAX:  (508) 872-5415
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-762-500-27

  Query Match             16.9%;  Score 257;  DB 3;  Length 1457;
  Best Local Similarity   24.9%;  Pred. No. 8.6e-18;
  Matches   91;  Conservative   40;  Mismatches   98;  Indels  136;  Gaps   15;

Qy       2 GKYPSLELQPWMYNEQYT--------FVSNDAPE-------DTGTLELLNALTKDPGFGT 46
           |  | | | |  |:  ||        : : :  |       |    :|::      | |
Db     504 GDLPPLVLSPSQYH-NYTQPRGNFIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGA 562

Qy      47 RCM-----------------------------------EGNPI--------------- 54
            |:                                   :| |:
Db     563 TCVLKSPANGSLGPMLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDS 622

Qy      55 ---PD----------TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKI 100
              ||           |  || | ||:|| :|: : :  :            | ||:
Db     623 PVXPDEDSLQAWNMSLPPTAGPETWTSAPSLPRLVHEPVR-----------CTCSAQGT 670

Qy     101 KKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEF 160
               ||   ||  ||| :  | ||| |:|| |:|:||: |  :
Db     671 GFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----------RLH 715

Qy     161 RYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
           |||  : |  | |   |:                    |   :   :  :  |
Db     716 RYGAITFG--NVQKSIPAS---------------------FGARVPPMVRKIAVRRVA 750

Qy     221 KVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQLS-EVAL 278
           :| :||||:|:: ::|| :|||||||||| | : ||: | ||  |||:| |   || :  |
Db     751 QVLYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYXITVTNHPMNKTSASLSLDYLL 810

Qy     279 MTTSV 283
             | |
Db     811 QGTDV 815


RESULT 5
5208144-8
;Patent No. 5208144
;    APPLICANT: SMITH, JOHN A.;RAYCHOWDHURY, RAKTIMA;NILES,JOHN L.
;    TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
;CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
;    NUMBER OF SEQUENCES: 42
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/396,697
;      FILING DATE: 22-AUG-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 313,682
;      FILING DATE: 22-FEB-1989
;      APPLICATION NUMBER: 235,211
;      FILING DATE: 23-AUG-1988
;SEQ ID NO:8:
;      LENGTH: 884
5208144-8

  Query Match              6.2%;  Score 94.5;  DB 6;  Length 884;
  Best Local Similarity   23.6%;  Pred. No. 0.57;
  Matches   70;  Conservative   37;  Mismatches  111;  Indels   79;  Gaps   19;

Qy       1 FGKYPS---LELQPWM----------YNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTR 47
           |||       | : ||:          ||:    |||   :    | ||    :  |:
Db     551 FGKENKEKVLVVNPWLTQVRIFHQLRYNQS---VSNPCKQVCSHLCLL----RPGGYSCA 603

Qy      48 CMEGNPI---PDTPCQAGEEEWTTAPVPQTIMDLFQNGN-WTMQNPSPACQCSSDKIKKM 103
           | :|:         | |  |   | | |   |     || :  :|  | |:|||    :
Db     604 CPQGSDFVTGSTVQCDAASELPVTMPPPCRCM---HGGNCYFDENELPKCKCSSGYSGE- 659
```

```
Qy    104 LPVCPPG-AGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRY 162
Db    660 --YCEVGLSRGIPP--------------GTTMA-VLLTFVIVIIVGAL---VLVGLFHY 698

Qy    163 GGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLG-RFMTGLDTRNNVK 221
Db    699 -------RKTGSLLPT------LPKLPSLSSLAKPSE-----NGNGVTFRSGADV--NMD 738

Qy    222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAL 278
Db    739 IGVSPFGPETIIDRSMAMNEHFV---MEVGKQP-----VIFENPMYAAKDNTSKVAL 787


RESULT 6
US-09-261-907-2
; Sequence 2, Application US/09261907A
; Patent No. 6294364
; GENERAL INFORMATION:
;  APPLICANT: ELLIS, CATHERINE
;  APPLICANT: LONSDALE, JOHN
;  APPLICANT: BERGSMA, DERK J.
;  APPLICANT: MOONEY, JEFFREY L.
;  APPLICANT: DePIERA, MEGAN E.
;  APPLICANT: CHAPMAN, CONRAD
;  TITLE OF INVENTION: HUMAN FAS
;  FILE REFERENCE: GP-70603
;  CURRENT APPLICATION NUMBER: US/09/261,907A
;  CURRENT FILING DATE: 1999-03-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2511
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-261-907-2

  Query Match             6.0%;  Score 92;  DB 4;  Length 2511;
  Best Local Similarity   23.2%;  Pred. No. 5.3;
  Matches   39;  Conservative   27;  Mismatches   60;  Indels   42;  Gaps    7;

Qy    107 CPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFS 166
Db    634 CPPGV--VPACHNSKDTVTI----SGPQAPVF------EFVEQLRKEGVFAKEVRTGGMA 681

Qy    167 LGVSNTQALPPS--QEVNDAIKQMKK----------HLKLAKDSSADRFLNSLGR 209
Db    682 FHSYFMEAIAPPLLQELKKVIREPKPRSARWLSTSIPEAQWHSSLARTSSAEYNVNNLVS 741

Qy    210 FMTGLDTRNNVKVWFNNKGWHA--ISSFLNVINNAILRANLQKGENPS 255
Db    742 ----------PVLFQEALWHVPEHAVVLEIAPHALLQAVLKRGLKPS 778


RESULT 7
US-08-481-190-8
; Sequence 8, Application US/08481190
; Patent No. 6160204
;  GENERAL INFORMATION:
;    APPLICANT:  John C. Steffens
;    TITLE OF INVENTION:  Polyphenol Oxidase cDNA
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Yahwak & Associates
;      STREET:  25 Skytop Drive
;      CITY:  Trumbull
;      STATE:  Connecticutt
;      COUNTRY:  USA
;      ZIP:  06611
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,190
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  203,533
;      FILING DATE:  02-24-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  George M. Yahwak
;      REGISTRATION NUMBER:  26,824
;      REFERENCE/DOCKET NUMBER:  UA 816 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203)268-1951
;      TELEFAX:  (203)268-1951
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  596 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-481-190-8

  Query Match             5.9%;  Score 89.5;  DB 4;  Length 596;
  Best Local Similarity   23.0%;  Pred. No. 1;
  Matches   63;  Conservative   32;  Mismatches  128;  Indels   51;  Gaps   12;

Qy     13 MYNEQYT-----------FVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db    293 MYRQMVTNAPCPLLFFGAPYVLGNNVEAPGTIETIPHIPVHIWAGT--VRGSKFPNGDVS 350

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
Db    351 YGED---------------MGNFYSAGLDPVFYCHHGNVDRMWNEW-KAIGG---KRRD 390

Qy    121 QNTADILQD---LTGRNISDYLVKTYVQIIAKSLKNKI------WVNEFRYGGFSLGVSN 171
Db    391 ISEKDWLNSEFFFYDEHKNPYRVKVRDCLDTKKMGYDYAPMPTPWRNFKPKSKASVGKVN 450

Qy    172 TQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHA 231
Db    451 TSTLPPANEVFPLAK-MDKTISFAINRPASSRTQQEKNEQEEMLTFNNIR--YDNRGYIR 507

Qy    232 ISSFLNVINNAILRAN-LQKGENPSHYGITAFNH 264
Db    508 FDVFLNVDNN--VNANELDKAEFAGSY--TSLPH 537


RESULT 8
PCT-US93-00869-8
; Sequence 8, Application PC/TUS9300869
;  GENERAL INFORMATION:
;    APPLICANT:  John C. Steffens
;    TITLE OF INVENTION:  Polyphenol Oxidase cDNAs: Cloning
;    TITLE OF INVENTION:  and Applications
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Yahwak & Associates
;      STREET:  25 Skytop Drive
;      CITY:  Trumbull
;      STATE:  Connecticut
;      COUNTRY:  USA
;      ZIP:  06611
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/00869
;      FILING DATE:  19930129
;      CLASSIFICATION:
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  George M. Yahwak
;      REGISTRATION NUMBER:  26,824
;      REFERENCE/DOCKET NUMBER:  CRF D-1057
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203)268-1951
;      TELEFAX:  (203)268-1951
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  596 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-00869-8

  Query Match             5.9%;  Score 89.5;  DB 5;  Length 596;
  Best Local Similarity   23.0%;  Pred. No. 1;
  Matches   63;  Conservative   32;  Mismatches  128;  Indels   51;  Gaps   12;

Qy      13 MYNEQYT-----------FVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           || : |            :|  :  |  ||:| :  :      ||  : |:  |:
Db     293 MYRQMVTNAPCPLLFFGAPYVLGNNVEAPGTIETIPHIPVHIWAGT--VRGSKFPNGDVS 350

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ||:              ||:      |   |    : :|      ||    :|
Db     351 YGED---------------MGNFYSAGLDPVFYCHHGNVDRMWNEW-KAIGG---KRRD 390

Qy     121 QNTADILQD---LTGRNISDYLVKTYVQIIAKSLKNKI------WVNEFRYGGFSLGVSN 171
            :  | |       : : | ||    :  | :           | |       |:|  |
Db     391 ISEKDWLNSEFFFYDEHKNPYRVKVRDCLDTKKMGYDYAPMPTPWRNFKPKSKASVGKVN 450

Qy     172 TQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHA 231
           |  |||: ||    | | | :  | :  |            : | ||::  ::|:|:
Db     451 TSTLPPANEVFPLAK-MDKTISFAINRPASSRTQQEKNEQEEMLTFNNIR--YDNRGYIR 507

Qy     232 ISSFLNVINNAILRAN-LQKGENPSHYGITAFNH 264
             |||| ||  : || | | |    |  |:  |
Db     508 FDVFLNVDNN--VNANELDKAEFAGSY--TSLPH 537


RESULT 9
US-08-804-439A-22
; Sequence 22, Application US/08804439A
; Patent No. 6015565
;  GENERAL INFORMATION:
;    APPLICANT:  Rose, Timothy M.
;    APPLICANT:  Bosch, Marnix L.
;    APPLICANT:  Strand, Kurt
;    TITLE OF INVENTION:  GLYCOPROTEIN B OF THE RFHV/KSHV
;    TITLE OF INVENTION:  SUBFAMILY OF HERPES VIRUSES
;    NUMBER OF SEQUENCES:  113
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Ste 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,439A
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09176/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-804-439A-22

  Query Match             5.8%;  Score 89;  DB 3;  Length 903;
  Best Local Similarity   20.2%;  Pred. No. 2.2;
  Matches   57;  Conservative   28;  Mismatches  115;  Indels   82;  Gaps   11;

Qy      21 VSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQ 80
           |:: ||   ||           ||        |  | ||          || |    |
Db      26 VASAAPSSPGT----------PGVAAATQAANGGPATPA-------PPAPGPAPTGDTKP 68

Qy      81 NGNWTMQNPSP--------------------ACQCSSDKIKKMLPVCPPGAGG------- 113
             |   :|| |                      ::       ||||  |
Db      69 KKNKKPKNPPPPRPAGDNATVAAGHATLREHLRDIKAENTDANFYVCPPPTGATVVQFEQ 128

Qy     114 ---LPPPQRKQNTADILQDLTGRNISDYLVKT--YVQIIAKSLKNKIWVNEFRYGGFSLG 168
               |    ||  :  :   :  ||: |  |   | : :  |   ::|    ||  | :|
Db     129 PRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYYKDVTVS---QVWFGH-RYSQF-MG 183

Qy     169 VSNTQALPPSQEVNDAI----------KQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRN 218
           :   :|  | :|| | |           | :: :|:      |          | :: :
Db     184 IFEDRAPVPFEEVIDKINAKGVCRSTAKYVRNNLETTAFHRDDH--------ETDMELKP 235

Qy     219 NVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGIT 260
                 ::|||            |: |  : |    || |
Db     236 ANAATRTSRGWHTTD----------LKYNPSRVEAFHRYGTT 267


RESULT 10
US-08-720-229-22
; Sequence 22, Application US/08720229
; Patent No. 6022542
;  GENERAL INFORMATION:
;    APPLICANT:  Rose, Timothy M.
;    APPLICANT:  Bosch, Marnix L.
;    APPLICANT:  Strand, Kurt
;    TITLE OF INVENTION:  GLYCOPROTEIN B OF THE RFHV/KSHV
;    TITLE OF INVENTION:  SUBFAMILY OF HERPES VIRUSES
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/720,229
;      FILING DATE:  26-SEP-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schiff, J. Michael
;      REGISTRATION NUMBER:  40,253
;      REFERENCE/DOCKET NUMBER:  29938-20002.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  22:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-720-229-22

  Query Match             5.8%;  Score 89;  DB 3;  Length 903;
  Best Local Similarity   20.2%;  Pred. No. 2.2;
  Matches   57;  Conservative   28;  Mismatches  115;  Indels   82;  Gaps   11;

Qy      21 VSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQ 80
           |:: ||   ||          ||        | | ||          || |    |
Db      26 VASAAPSSPGT----------PGVAAATQAANGGPATPA-------PPAPGPAPTGDTKP 68

Qy      81 NGNWTMQNPSP--------------------ACQCSSDKIKKMLPVCPPGAGG------- 113
             |   :|| |                    ::       ||||  |
Db      69 KKNKKPKNPPPPRPAGDNATVAAGHATLREHLRDIKAENTDANFYVCPPPTGATVVQFEQ 128

Qy     114 ---LPPPQRKQNTADILQDLTGRNISDYLVKT--YVQIIAKSLKNKIWVNEFRYGGFSLG 168
               |     ||  : :  :   ||: | |   | : :  |   ::|    ||  | :|
Db     129 PRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYYKDVTVS---QVWFGH-RYSQF-MG 183

Qy     169 VSNTQALPPSQEVNDAI----------KQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRN 218
           ;   :|  | :|| | |          | :: :|:       |          | :: :
Db     184 IFEDRAPVPFEEVIDKINAKGVCRSTAKYVRNNLETTAFHRDDH--------ETDMELKP 235

Qy     219 NVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGIT 260
                 ::|||           |: |  : |    || |
Db     236 ANAATRTSRGWHTTD---------LKYNPSRVEAFHRYGTT 267


RESULT 11
US-08-220-151-8
; Sequence 8, Application US/08220151
; Patent No. 5529780
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Limbach, Keith J.
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF
;    TITLE OF INVENTION:  CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/220,151
;      FILING DATE:  30-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2540
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-220-151-8

  Query Match             5.8%;  Score 88.5;  DB 1;  Length 903;
  Best Local Similarity   20.6%;  Pred. No. 2.5;
  Matches   58;  Conservative   32;  Mismatches  109;  Indels   83;  Gaps   13;

Qy      21 VSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQ 80
           |:: ||   ||       : :||| | |          || :|      ||      |
Db      27 VASAAPSSPGT----PGVARDPG-GER---------GPCHSGAAALGAAPTG----DPKP 68

Qy      81 NGNWTMQNPSP--------------------ACQCSSDKIKKMLPVCPPGAGG------- 113
             |   :||:|                    ::       ||||  |
Db      69 KKNKKPKNPTPPRPAGDNATVAAGHATLREHLRDIKAENTDANFYVCPPPTGATVVQFEQ 128

Qy     114 ---LPPPQRKQNTADILQDLTGRNISDYLVKT--YVQIIAKSLKNKIWVNEFRYGGFSLG 168
               |     ||  : :  :   ||: | |   | : :  |   ::|    ||  | :|
Db     129 PRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYYKDVTVS---QVWFGH-RYSQF-MG 183

Qy     169 VSNTQALPPSQEVNDAI----------KQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRN 218
           :   :|  | :|| | |          | :: :|:       |          | :: :
Db     184 IFEDRAPVPFEEVIDKINAKGVCRSTAKYVRNNLETTAFHRDDH--------ETDMELKP 235

Qy     219 NVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGIT 260
                 ::|||           |: |  : |    || |
Db     236 ANAATRTSRGWHTTD----------LKYNPSRVEAFHRYGTT 267


RESULT 12
US-08-413-118-8
; Sequence 8, Application US/08413118
; Patent No. 5688920
;  GENERAL INFORMATION:
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  LIMBACH, KEITH J.
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF
;    TITLE OF INVENTION:  CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CURTIS, MORRIS & SAFFORD, P.C.
;      STREET:  530 FIFTH AVENUE, 25TH FLOOR
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/413,118
;      FILING DATE:  29-MAR-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/220,151
;      FILING DATE:  30-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FROMMER, WILLIAM S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2670
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
```

```
US-08-413-118-8

  Query Match             5.8%;  Score 88.5;  DB 1;  Length 903;
  Best Local Similarity   20.6%;  Pred. No. 2.5;
  Matches   58;  Conservative   32;  Mismatches  109;  Indels   83;  Gaps   13;

Qy      21 VSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQ 80
           |:: ||   ||      : :||| | |          || :|      ||      |
Db      27 VASAAPSSPGT----PGVARDPG-GER---------GPCHSGAAALGAAPTG----DPKP 68

Qy      81 NGNWTMQNPSP-------------------ACQCSSDKIKKMLPVCPPGAGG------- 113
             |   :||:|                    ::       ||||  |
Db      69 KKNKKPKNPTPPRPAGDNATVAAGHATLREHLRDIKAENTDANFYVCPPPTGATVVQFEQ 128

Qy     114 ---LPPPQRKQNTADILQDLTGRNISDYLVKT--YVQIIAKSLKNKIWVNEFRYGGFSLG 168
               |     ||  : :  :   ||: |  |   | : :  |   ::|    ||  | :|
Db     129 PRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYYKDVTVS---QVWFGH-RYSQF-MG 183

Qy     169 VSNTQALPPSQEVNDAI----------KQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRN 218
           :   :|  | :|| | |          | :: :|:       |          | :: :
Db     184 IFEDRAPVPFEEVIDKINAKGVCRSTAKYVRNNLETTAFHRDDH--------ETDMELKP 235

Qy     219 NVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGIT 260
               ::|||             |: |  : |    || |
Db     236 ANAATRTSRGWHTTD----------LKYNPSRVEAFHRYGTT 267


RESULT 13
US-08-473-446-8
; Sequence 8, Application US/08473446
; Patent No. 6017542
;  GENERAL INFORMATION:
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  LIMBACH, KEITH J.
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF
;    TITLE OF INVENTION:  CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CURTIS, MORRIS & SAFFORD, P.C.
;      STREET:  530 FIFTH AVENUE, 25TH FLOOR
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,446
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/413,118
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FROMMER, WILLIAM S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2670
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  903 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-473-446-8

  Query Match             5.8%;  Score 88.5;  DB 3;  Length 903;
  Best Local Similarity   20.6%;  Pred. No. 2.5;
  Matches   58;  Conservative   32;  Mismatches  109;  Indels   83;  Gaps   13;

Qy      21 VSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQ 80
           |:: ||   ||      : :||| | |          || :|      ||      |
Db      27 VASAAPSSPGT----PGVARDPG-GER---------GPCHSGAAALGAAPTG----DPKP 68

Qy      81 NGNWTMQNPSP-------------------ACQCSSDKIKKMLPVCPPGAGG------- 113
             |   :||:|                    ::       ||||  |
Db      69 KKNKKPKNPTPPRPAGDNATVAAGHATLREHLRDIKAENTDANFYVCPPPTGATVVQFEQ 128

Qy     114 ---LPPPQRKQNTADILQDLTGRNISDYLVKT--YVQIIAKSLKNKIWVNEFRYGGFSLG 168
               |     ||  : :  :   ||: |  |   | : :  |   ::|    ||  | :|
Db     129 PRRCPTRPEGQNYTEGIAVVFKENIAPYKFKATMYYKDVTVS---QVWFGH-RYSQF-MG 183

Qy     169 VSNTQALPPSQEVNDAI----------KQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRN 218
           :   :|  | :|| | |          | :: :|:       |          | :: :
Db     184 IFEDRAPVPFEEVIDKINAKGVCRSTAKYVRNNLETTAFHRDDH--------ETDMELKP 235

Qy     219 NVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGIT 260
               ::|||             |: |  : |    || |
Db     236 ANAATRTSRGWHTTD----------LKYNPSRVEAFHRYGTT 267


RESULT 14
US-08-861-464-6
; Sequence 6, Application US/08861464
; Patent No. 5874210
;  GENERAL INFORMATION:
;    APPLICANT:  Guarente, Leonard P.
;    APPLICANT:  Austriaco Jr., Nicanor
;    APPLICANT:  Kennedy, Brian
;    TITLE OF INVENTION:  Genes Determining Cellular Senescence
;    TITLE OF INVENTION:  in Yeast
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/861,464
;      FILING DATE:  22-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/396,001
;      FILING DATE:  28-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/09351
;      FILING DATE:  15-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,408
;      FILING DATE:  16-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-6408A2Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  781-861-6240
;      TELEFAX:  781-861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  888 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-861-464-6

  Query Match             5.8%;  Score 88;  DB 2;  Length 888;
  Best Local Similarity   19.4%;  Pred. No. 2.8;
  Matches   61;  Conservative   36;  Mismatches  105;  Indels  112;  Gaps   13;

Qy      24 DAPEDTGTLE-LLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDL-FQN 81
           | ||:  |     | :::|  |             | |         | || :  : |
Db      34 DDPEENATSNAFANKVSQDSQFAN---------GPPSQ-------MFPHPQMMGGMGFMP 77

Qy      82 GNWTMQNPSPACQCSSDKIKKMLPVCPP--------------GAGGLPPPQRKQNTADI 126
             :  || |   |           |  ||                |||  ||   :| :
Db      78 YSQMMQVPHNPC-----------PFFPPPDFNDPTAPLSSSPLNAGG--PPMLFKNDSLP 124

Qy     127 LQDLT---------GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPP 177
           | |:         |:|::  :    :::  :  : :| :    |  |: :  | |
Db     125 FQMLSSGAAVATQGGQNLNPLINDNSMKVLPIASADPLWTHSNVPGSASVAIEETTA--- 181

Qy     178 SQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAIS--SF 235
                       :   | | ||       ||    |  : :||:|
Db     182 --------------------TLQESLPSKGR------ESNNKASSFRRQTFHALSPTDL 214

Qy     236 LNVINNAIL-------RANLQKGENPS-----------------HYGITAFNHPLNLTK 270
           :|  ||  |         |  | : ||                   |:|  | |
Db     215 INAANNVTLSKDFQSDMQNFSKAKKPSVGANNTAKTRTQSISFDNTPSSTSFIPPTNSVS 274

Qy     271 QQLSEVALMTTSVD 284
           ::||:  : |:  |
Db     275 EKLSDFKIETSKED 288


RESULT 15
US-08-396-001-6
; Sequence 6, Application US/08396001
; Patent No. 5919618
;  GENERAL INFORMATION:
;    APPLICANT:  Guarente, Leonard P.
;    APPLICANT:  Austriaco Jr., Nicanor
;    APPLICANT:  Claus, James
;    APPLICANT:  Cole, Francesca
;    APPLICANT:  Kennedy, Brian
;    TITLE OF INVENTION:  Genes Determining Cellular Senesence in
;    TITLE OF INVENTION:  Yeast
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,001
;      FILING DATE:  28-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-6408A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  888 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear

;    MOLECULE TYPE:  protein
US-08-396-001-6

  Query Match             5.8%;  Score 88;  DB 2;  Length 888;
  Best Local Similarity   19.4%;  Pred. No. 2.8;
  Matches   61;  Conservative   36;  Mismatches  105;  Indels  112;  Gaps   13;

Qy      24 DAPEDTGTLE-LLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDL-FQN 81
           | ||:  |     | :::|  |             | |         | || :  : |
Db      34 DDPEENATSNAFANKVSQDSQFAN---------GPPSQ-------MFPHPQMMGGMGFMP 77

Qy      82 GNWTMQNPSPACQCSSDKIKKMLPVCPP--------------GAGGLPPPQRKQNTADI 126
             :  || |   |           |  ||                |||  ||   :| :
Db      78 YSQMMQVPHNPC-----------PFFPPPDFNDPTAPLSSSPLNAGG--PPMLFKNDSLP 124

Qy     127 LQDLT---------GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPP 177
           | |:         |:|::  :    :::  :  : :| :    |  |: :  | |
Db     125 FQMLSSGAAVATQGGQNLNPLINDNSMKVLPIASADPLWTHSNVPGSASVAIEETTA--- 181

Qy     178 SQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAIS--SF 235
                       :   | | ||       ||    |  : :||:|
Db     182 --------------------TLQESLPSKGR------ESNNKASSFRRQTFHALSPTDL 214

Qy     236 LNVINNAIL-------RANLQKGENPS-----------------HYGITAFNHPLNLTK 270
           :|  ||  |         |  | : ||                   |:|  | |
Db     215 INAANNVTLSKDFQSDMQNFSKAKKPSVGANNTAKTRTQSISFDNTPSSTSFIPPTNSVS 274

Qy     271 QQLSEVALMTTSVD 284
           ::||:  : |:  |
Db     275 EKLSDFKIETSKED 288


Search completed: February  4, 2003, 09:43:28
Job time : 19 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:40:23 ; Search time 13 Seconds
(without alignments)
484.314 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 129505 seqs, 22169297 residues

Total number of hits satisfying chosen parameters: 129505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1513	99.2	2261	10	US-09-995-542-11	Sequence 11, Appl
2	1513	99.2	2261	10	US-09-846-456-11	Sequence 11, Appl
3	1423	93.3	2201	10	US-09-995-542-9	Sequence 9, Appli
4	733.5	48.1	2273	10	US-09-995-542-12	Sequence 12, Appl
5	724.5	47.5	2310	10	US-09-995-542-10	Sequence 10, Appl
6	664	43.5	2121	10	US-09-995-542-3	Sequence 3, Appli
7	664	43.5	2167	10	US-09-995-542-2	Sequence 2, Appli
8	663.5	43.5	1550	10	US-09-995-542-8	Sequence 8, Appli
9	663.5	43.5	2100	10	US-09-995-542-6	Sequence 6, Appli
10	663.5	43.5	2146	10	US-09-995-542-5	Sequence 5, Appli
11	662.5	43.4	2144	10	US-09-858-194-2	Sequence 2, Appli
12	460.5	30.2	199	10	US-09-767-870-18	Sequence 18, Appl
13	267	17.5	2001	9	US-10-072-621-8	Sequence 8, Appli
14	267	17.5	2436	10	US-09-795-693-8	Sequence 8, Appli
15	140	9.2	664	10	US-09-767-870-9	Sequence 9, Appli
16	88	5.8	888	10	US-09-826-752-6	Sequence 6, Appli
17	87.5	5.7	522	10	US-09-876-889-353	Sequence 353, App
18	86.5	5.7	969	9	US-09-981-353-122	Sequence 122, App
19	86.5	5.7	977	10	US-09-925-297-797	Sequence 797, App
20	85.5	5.6	953	9	US-10-118-984-43	Sequence 43, Appl
21	85.5	5.6	953	10	US-09-728-721-43	Sequence 43, Appl
22	82.5	5.4	384	9	US-10-029-180-50	Sequence 50, Appl
23	81.5	5.3	172	10	US-09-764-847-551	Sequence 551, App
24	81.5	5.3	774	10	US-09-815-242-12046	Sequence 12046, A
25	81.5	5.3	972	10	US-09-924-154-16	Sequence 16, Appl
26	80	5.2	426	9	US-09-909-650A-24	Sequence 24, Appl
27	80	5.2	426	9	US-09-165-522-10	Sequence 10, Appl
28	80	5.2	464	9	US-09-165-522-2	Sequence 2, Appli
29	80	5.2	678	9	US-09-895-913A-4	Sequence 4, Appli
30	79.5	5.2	1242	10	US-09-903-248-5	Sequence 5, Appli
31	79.5	5.2	1242	10	US-09-859-604-5	Sequence 5, Appli
32	79.5	5.2	1242	10	US-09-903-063-5	Sequence 5, Appli
33	79.5	5.2	1242	10	US-09-903-216-5	Sequence 5, Appli
34	79.5	5.2	1242	10	US-09-903-199-5	Sequence 5, Appli
35	79.5	5.2	1242	10	US-09-903-023-5	Sequence 5, Appli
36	79.5	5.2	1242	12	US-10-085-027-1	Sequence 1, Appli
37	79	5.2	1848	9	US-09-839-996-6	Sequence 6, Appli
38	78.5	5.1	793	10	US-09-881-752A-362	Sequence 362, App
39	78.5	5.1	2434	10	US-09-815-242-5835	Sequence 5835, Ap
40	78.5	5.1	6281	10	US-09-815-242-12996	Sequence 12996, A
41	78	5.1	704	10	US-09-801-368-218	Sequence 218, App
42	77	5.0	315	10	US-09-764-853-793	Sequence 793, App
43	76.5	5.0	364	10	US-09-756-983-22	Sequence 22, Appl
44	76.5	5.0	378	10	US-09-801-574-38	Sequence 38, Appl
45	76.5	5.0	540	9	US-09-738-626-4450	Sequence 4450, Ap

ALIGNMENTS

RESULT 1
US-09-995-542-11
; Sequence 11, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
; APPLICANT: Shutter, John
; APPLICANT: Ulias, Laarni
; TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-995-542-11

Query Match 99.2%; Score 1513; DB 10; Length 2261;
Best Local Similarity 99.3%; Pred. No. 2.6e-138;
Matches 282; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDATKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610
```

```
Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||||||||||||||||||||||||||||||||||| ||||||
Db   1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAPMTTSVD 1654


RESULT 2
US-09-846-456-11
; Sequence 11, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
;  APPLICANT: Rosier, Marie
;  APPLICANT:  Prades, Catherine
;  APPLICANT:  Lemoine, Cendrine
;  APPLICANT:  Naudin, Laurent
;  APPLICANT:  Denefle, Patrice
;  APPLICANT:  Duverger, Nicolas
;  APPLICANT:  Brewer, Bryan
;  APPLICANT:  Remaley, Alan
;  APPLICANT:  Fojo, Silvia
;  TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
;  TITLE OF INVENTION:  Activity and Therapeutic Uses
;  FILE REFERENCE: 3806.0505
;  CURRENT APPLICATION NUMBER: US/09/846,456
;  CURRENT FILING DATE:  2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/201,280
;  PRIOR FILING DATE: 2000-05-02
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 2261
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-456-11

  Query Match             99.2%;  Score 1513;  DB 10;  Length 2261;
  Best Local Similarity   99.3%;  Pred. No. 2.6e-138;
  Matches  282;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 VNDATKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||||||||||||||||||||||||||||||||||| ||||||
Db   1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAPMTTSVD 1654


RESULT 3
US-09-995-542-9
; Sequence 9, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 2201
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (115)
;   OTHER INFORMATION: amino acid at this position is unknown
US-09-995-542-9

  Query Match             93.3%;  Score 1423;  DB 10;  Length 2201;
  Best Local Similarity   93.0%;  Pred. No. 1.4e-129;
  Matches  264;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          |||||||||||||||||||||||||||| || ||||||||||||||||||||||||||||
Db   1311 FGKYPSLELQPWMYNEQYTFVSNDAPEDMGTQELLNALTKDPGFGTRCMEGNPIPDTPCL 1370

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          ||||:|| :||||:|:||||||||||:||||||||||||||||||||||||||||||||||
Db   1371 AGEEDWTISPVPQSIVDLFQNGNWTMKNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1430

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | ||||||:||||||||||||||||||||||||||||||||||||||||:|||||| |
Db   1431 QKTADILQNLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHE 1490

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          |||||||||| ||| ||:||||||:|||||| ||||:||||||||||||||||||||||||
Db   1491 VNDAIKQMKKLLKLTKDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLNVIN 1550

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||||||||||| ||||||||||||||||||||||||||||
Db   1551 NAILRANLQKGENPSQYGITAFNHPLNLTKQQLSEVALMTTSVD 1594


RESULT 4
US-09-995-542-12
; Sequence 12, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2273
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-995-542-12

  Query Match             48.1%;  Score 733.5;  DB 10;  Length 2273;
  Best Local Similarity   48.2%;  Pred. No. 2e-62;
  Matches  146;  Conservative   35;  Mismatches   83;  Indels   39;  Gaps    4;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||:||:| | ||:| :|||| | | |       | : |   |||| ||::   :|: ||
Db   1397 FGEYPALTLHPWIYGQQYTFFSMDEPGSEQFTVLADVLLNKPGFGNRCLKEGWLPEYPC- 1455

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
              | |   |   | |||   ||  ||||:|:||: :   ||| || ||||||||||
Db   1456 GNSTPWKTPSVSPNITQLFQKQKWTQVNPSPSCRCSTREKLTMLPECPEGAGGLPPPQRT 1515

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | : :|||||| |||||:|||||  :|  |||:| |||| |||| |:|      ||
```

```
Db   1516 QRSTEILQDLTDRNISDFLVKTYPALIRSSLKSKFWVNEQRYGGISIG----GKLPVVPI 1571

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGR------------------FMTGLDTRNNVK 221
Db   1572 TGEALV--------------GFLSDLGRIMNVSGGPITREASKEIPDFLKHLETEDNIK 1616

Qy    222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTT 281
Db   1617 VWFNNKGWHALVSFLNVAHNAILRASLPKDRSPEEYGITVISQPLNLTKEQLSEITVLTT 1676

Qy    282 SVD 284
Db   1677 SVD 1679


RESULT 5
US-09-995-542-10
; Sequence 10, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 2310
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-995-542-10

  Query Match             47.5%;  Score 724.5;  DB 10;  Length 2310;
  Best Local Similarity   47.2%;  Pred. No. 1.5e-61;
  Matches  143;  Conservative   39;  Mismatches   82;  Indels   39;  Gaps    4;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db   1396 FGEFPALTLHPWMYGHQYTFFSMDEPNNEHLEVLADVLLNRPGFGNRCLKEEWLPEYPC- 1454

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
Db   1455 INATSWKTPSVSPNITHLFQKQKWTAAHPSPSCKCSTREKLTMLPECPEGAGGLPPPQRT 1514

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
Db   1515 QRSTEVLQDLTNRNISDYLVKTYPALIRSSLKSKFWVNEQRYGGISIG-GKLPAIPISGE 1573

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGR------------------FMTGLDTRNNVK 221
Db   1574 ------------------ALVGFLSGLGQMMNVSGGPVTREASKEMLDFLKHLETTDNIK 1615

Qy    222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTT 281
Db   1616 VWFNNKGWHALVSFLNVAHNAILRASLPRDRDPEEYGITVISQPLNLTKEQLSDITVLTT 1675

Qy    282 SVD 284
Db   1676 SVD 1678


RESULT 6
US-09-995-542-3
; Sequence 3, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2121
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-995-542-3

  Query Match             43.5%;  Score 664;  DB 10;  Length 2121;
  Best Local Similarity   45.3%;  Pred. No. 1e-55;
  Matches  129;  Conservative   47;  Mismatches  107;  Indels    2;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db   1228 FGQYPPLQLSPAMYGPQVSFFSEDAPGDPNRMKLLEALLGEAGLQEPSMQDKDARGSECT 1287

Qy     61 AGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQR 119
Db   1288 HSLACYFTVPEVPPDVASILASGNWTPESPSPACQCSQPGARRLLPDCPAGAGGPPPPQA 1347

Qy    120 KQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQ 179
Db   1348 VAGLGEVVQNLTGRNVSDFLVKTYPSLVRRGLKTKKWVDEVRYGGFSLG-GRDPDLPTGH 1406

Qy    180 EVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVI 239
Db   1407 EVVRTLAEIRALLSPQPGNALDRILNNLTQWALGLDARNSLKIWFNNKGWHAMVAFVNRA 1466

Qy    240 NNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
Db   1467 NNGLLHALLPSGPVRHAHSITTLNHPLNLTKEQLSEATLIASSVD 1511


RESULT 7
US-09-995-542-2
; Sequence 2, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2167
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-995-542-2

  Query Match             43.5%;  Score 664;  DB 10;  Length 2167;
  Best Local Similarity   45.3%;  Pred. No. 1.1e-55;
  Matches  129;  Conservative   47;  Mismatches  107;  Indels    2;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
Db   1274 FGQYPPLQLSPAMYGPQVSFFSEDAPGDPNRMKLLEALLGEAGLQEPSMQDKDARGSECT 1333

Qy     61 AGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQR 119
Db   1334 HSLACYFTVPEVPPDVASILASGNWTPESPSPACQCSQPGARRLLPDCPAGAGGPPPPQA 1393
```

```
Qy    120 KQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQ 179
          :::|:|||||:||:|||||  :: : || | ||:| ||||||||       ||
Db   1394 VAGLGEVVQNLTGRNVSDFLVKTYPSLVRRGLKTKKWVDEVRYGGFSLG-GRDPDLPTGH 1452

Qy    180 EVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVI 239
          ||   : :::  |     :: || ||:| ::  ||| ||::|:||||||||: :|:|
Db   1453 EVVRTLAEIRALLSPQPGNALDRILNNLTQWALGLDARNSLKIWFNNKGWHAMVAFVNRA 1512

Qy    240 NNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          || :| | |  |     : ||  ||||||||:|||| |: :|||
Db   1513 NNGLLHALLPSGPVRHAHSITTLNHPLNLTKEQLSEATLIASSVD 1557


RESULT 8
US-09-995-542-8
; Sequence 8, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1550
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-995-542-8

  Query Match             43.5%;  Score 663.5;  DB 10;  Length 1550;
  Best Local Similarity   44.7%;  Pred. No. 7.4e-56;
  Matches  127;  Conservative   50;  Mismatches   96;  Indels   11;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db   1125 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1174

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  : :|||| ::|||||||    :::|| ||  ||| ||||
Db   1175 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSQPGARRLLPDCPAAAGGPPPPQAV 1234

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
            : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db   1235 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1293

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          :  :::::   |      : || | :|  :   || ::::|:|||||||:: :|:|  :
Db   1294 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1353

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||:|  |      : ||  |||||||:|||| ||| :|||
Db   1354 NAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEAALMASSVD 1397


RESULT 9
US-09-995-542-6
; Sequence 6, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 2100
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-995-542-6

  Query Match             43.5%;  Score 663.5;  DB 10;  Length 2100;
  Best Local Similarity   44.7%;  Pred. No. 1.2e-55;
  Matches  127;  Conservative   50;  Mismatches   96;  Indels   11;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db   1217 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1266

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  : :|||| ::|||||||    :::|| ||  ||| ||||
Db   1267 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSQPGARRLLPDCPAAAGGPPPPQAV 1326

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
            : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db   1327 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1385

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          :  :::::   |      : || | :|  :   || ::::|:|||||||:: :|:|  :
Db   1386 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1445

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||:|  |      : ||  |||||||:|||| ||| :|||
Db   1446 NAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEAALMASSVD 1489


RESULT 10
US-09-995-542-5
; Sequence 5, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
;  APPLICANT: Shutter, John
;  APPLICANT:  Ulias, Laarni
;  TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 00-658-A
;  CURRENT APPLICATION NUMBER: US/09/995,542
;  CURRENT FILING DATE:  2001-11-28
;  PRIOR APPLICATION NUMBER: 60/253,520
;  PRIOR FILING DATE: 2000-11-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 2146
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-995-542-5

  Query Match             43.5%;  Score 663.5;  DB 10;  Length 2146;
  Best Local Similarity   44.7%;  Pred. No. 1.2e-55;
  Matches  127;  Conservative   50;  Mismatches   96;  Indels   11;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db   1263 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1312

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  : :|||| ::|||||||    :::|| ||  ||| ||||
Db   1313 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSQPGARRLLPDCPAAAGGPPPPQAV 1372

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
            : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db   1373 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1431
```

```
Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          :  :::::   |      : || | :|  :   || ::::|:||||||||:: :|:|  :
Db   1432 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1491

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||:|  |      : ||  |||||||:|||| ||| :|||
Db   1492 NAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEAALMASSVD 1535


RESULT 11
US-09-858-194-2
; Sequence 2, Application US/09858194
; Patent No. US20020061590A1
; GENERAL INFORMATION:
;  APPLICANT: GLUCKSMANN, MARIA
;  APPLICANT:  CURTIS, RORY A.J.
;  TITLE OF INVENTION: 38594, A NOVEL HUMAN TRANSPORTER AND USES THEREOF
;  FILE REFERENCE: MNI-153
;  CURRENT APPLICATION NUMBER: US/09/858,194
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: 60/204,211
;  PRIOR FILING DATE: 2000-05-12
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2144
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-858-194-2

  Query Match             43.4%;  Score 662.5;  DB 10;  Length 2144;
  Best Local Similarity   44.7%;  Pred. No. 1.5e-55;
  Matches  127;  Conservative   50;  Mismatches   96;  Indels   11;  Gaps    2;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db   1261 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1310

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  :  :|||| ::|||||||||    :::|| ||  ||| ||||
Db   1311 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSRPGARRLLPDCPAAAGGPPPPQAV 1370

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
            : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db   1371 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1429

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          :  :::::   |      : || | :|  :   || ::::|:||||||||:: :|:|  :
Db   1430 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1489

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          ||||||:|  |      : ||  |||||||:|||| ||| :|||
Db   1490 NAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEAALMASSVD 1533


RESULT 12
US-09-767-870-18
; Sequence 18, Application US/09767870
; Patent No. US20020037549A1
; GENERAL INFORMATION:
;  APPLICANT:  Ruben et al.
;  TITLE OF INVENTION:  ABC Transport Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE:  PT010P1
;  CURRENT APPLICATION NUMBER: US/09/767,870
;  CURRENT FILING DATE:  2001-01-24
;  PRIOR APPLICATION NUMBER:  PCT/US00/19736
;  PRIOR FILING DATE:  2000-07-20
;  PRIOR APPLICATION NUMBER:  60/145,215
;  PRIOR FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:  60/149,445
;  PRIOR FILING DATE:  1999-08-18
;  PRIOR APPLICATION NUMBER:  60/164,730
;  PRIOR FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 199
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-767-870-18

  Query Match             30.2%;  Score 460.5;  DB 10;  Length 199;
  Best Local Similarity   41.0%;  Pred. No. 2.1e-37;
  Matches   86;  Conservative   40;  Mismatches   73;  Indels   11;  Gaps    2;

Qy     22 SNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQAGEEEWTTAPVPQTIMDLFQN 81
          | ||| | |   || || :: |          : : | |     ::   ||  :  :  :
Db      1 SEDAPGDPGRARLLEALLQEAG----------LEEPPVQHSSHRFSAPEVPAEVAKVLAS 50

Qy     82 GNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVK 141
          |||| ::|||||||    :::|| ||  ||| ||||    : :::|:|||||:||:|||
Db     51 GNWTPESPSPACQCSRPGARRLLPDCPAAAGGPPPPQAVTGSGEVVQNLTGRNLSDFLVK 110

Qy    142 TYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSAD 201
          || ::: : || | |||| ||||||||      ||  ||:  :::::   |      : |
Db    111 TYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQELGRSVEELWALLSPLPGGALD 169

Qy    202 RFLNSLGRFMTGLDTRNNVKVWFNNKGWHA 231
          | | :|  :   || ::::|:||||||||:
Db    170 RVLKNLTAWAHSLDAQDSLKIWFNNKGWHS 199


RESULT 13
US-10-072-621-8
; Sequence 8, Application US/10072621
; Patent No. US20020169137A1
; GENERAL INFORMATION:
;  APPLICANT: Reiner, Peter B.
;  APPLICANT:  Connop, Bruce P.
;  APPLICANT:  Pollard, Michelle
;  TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
;  TITLE OF INVENTION:  BY MODIFICATION OF ABC TRANSPORTER EXPRESSION OR ACTIVITY
;  FILE REFERENCE: 100103.402
;  CURRENT APPLICATION NUMBER: US/10/072,621
;  CURRENT FILING DATE:  2002-02-08
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 2001
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
;   OTHER INFORMATION: Xaa = Any Amino Acid
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-10-072-621-8

  Query Match             17.5%;  Score 267;  DB 9;  Length 2001;
  Best Local Similarity   28.8%;  Pred. No. 4.3e-17;
  Matches   79;  Conservative   41;  Mismatches   68;  Indels   86;  Gaps   13;

Qy     44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
          | : |:|                :| | |:|              ||:| ||:|| :
Db   1137 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGQEMWTSAPSL 1196

Qy     72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
          |: : :  :            | ||:         ||   ||  ||| :  | ||| |:|
Db   1197 PRLVREPVR-----------CTCSAQGTGFS---CPNSVGG-HPPQMRVVTGDILTDIT 1240

Qy    132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
          | |:|:||: |  :              |||  : | :  :::| |          | :
Db   1241 GHNVSEYLLFTSDRF-----------RLHRYGAITFG-NVLKSIPAS--FGTRAPPMVRK 1286
```

```
Qy    192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
          :: |: :                      :|::||||:|:: ::|| :|||||||||| |
Db   1287 IRCARAA---------------------QVFYNNKGYHSMPTYLNSLNNAILRANLPKS 1324

Qy    252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
          : ||: ||||  |||:| |   || :  |  | |
Db   1325 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 1358


RESULT 14
US-09-795-693-8
; Sequence 8, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
;  APPLICANT:  Glucksmann, Maria A.
;  TITLE OF INVENTION:  20685, 579, 17114, 23821, 33894, and
;  TITLE OF INVENTION:  32613, No. US20020068710A1el Human Transporters
;  FILE REFERENCE:  35800/209292
;  CURRENT APPLICATION NUMBER: US/09/795,693
;  CURRENT FILING DATE:  2001-02-28
;  PRIOR APPLICATION NUMBER:  60/185,906
;  PRIOR FILING DATE:  2000-02-29
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 2436
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-795-693-8

  Query Match             17.5%;  Score 267;  DB 10;  Length 2436;
  Best Local Similarity   29.2%;  Pred. No. 5.7e-17;
  Matches   80;  Conservative   35;  Mismatches   73;  Indels   86;  Gaps   12;

Qy     44 FGTRCMEG---------------NPIP-DTPCQ---------------AGEEEWTTAP-V 71
          | : |:|                :| | |:|                 || | ||:|| :
Db   1572 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 1631

Qy     72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
          |: : :  :             | ||:        ||   ||  ||| :  | ||| |:|
Db   1632 PRLVREPVR-----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDIT 1675

Qy    132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
          | |:|:||: |  :              |||  : |  |     |:     |   ::|
Db   1676 GHNVSEYLLFTSDRF-----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 1721

Qy    192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                              :  |   :|::||||:|:: ::|| :|||||||||| |
Db   1722 ---------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 1759

Qy    252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
          : ||: ||||  |||:| |   || :  |  | |
Db   1760 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 1793


RESULT 15
US-09-767-870-9
; Sequence 9, Application US/09767870
; Patent No. US20020037549A1
; GENERAL INFORMATION:
;  APPLICANT:  Ruben et al.
;  TITLE OF INVENTION:  ABC Transport Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE:  PT010P1
;  CURRENT APPLICATION NUMBER: US/09/767,870
;  CURRENT FILING DATE:  2001-01-24
;  PRIOR APPLICATION NUMBER:  PCT/US00/19736
;  PRIOR FILING DATE:  2000-07-20
;  PRIOR APPLICATION NUMBER:  60/145,215
;  PRIOR FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:  60/149,445
;  PRIOR FILING DATE:  1999-08-18
;  PRIOR APPLICATION NUMBER:  60/164,730
;  PRIOR FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 664
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-767-870-9

  Query Match              9.2%;  Score 140;  DB 10;  Length 664;
  Best Local Similarity   58.8%;  Pred. No. 2e-05;
  Matches   30;  Conservative    7;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    234 SFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          :|:|  :|||||:|  |     : ||  ||||||||:|||| ||| :|||
Db      3 AFVNRASNAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEGALMASSVD 53


Search completed: February  4, 2003, 09:44:20
Job time : 22 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:38:27 ; Search time 19 Seconds
(without alignments)
1436.957 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1423	93.3	2201	2	A54774	ATP binding casset
2	267	17.5	1529	2	A59189	ATP-binding casset
3	257	16.9	1472	2	B54774	ATP binding casset
4	215.5	14.1	1447	2	T15200	hypothetical prote
5	108.5	7.1	432	2	T14292	glutamate-ammonia
6	100.5	6.6	434	1	AJBHQ	glutamate-ammonia
7	98	6.4	877	2	F90070	Clumping factor B
8	97	6.4	908	2	T16057	hypothetical prote
9	95.5	6.3	429	1	AJFBQD	glutamate-ammonia
10	94.5	6.2	263	2	C64339	hypothetical prote
11	94.5	6.2	430	2	S18600	glutamate-ammonia
12	94.5	6.2	4660	2	T42737	gp330 protein prec
13	92.5	6.1	428	1	AJRZQD	glutamate-ammonia
14	91	6.0	459	2	B83793	hypothetical prote
15	90.5	5.9	428	2	S32228	glutamate-ammonia
16	90	5.9	773	2	F90537	lipoprotein [impor
17	89.5	5.9	423	2	S39482	glutamate-ammonia
18	89.5	5.9	596	1	S33540	catechol oxidase (
19	89.5	5.9	649	2	B96729	hypothetical prote
20	89	5.8	903	1	VGBEB1	glycoprotein B pre
21	88.5	5.8	363	2	S38154	hypothetical prote
22	88.5	5.8	865	2	AG2023	hypothetical prote
23	88.5	5.8	903	1	VGBEK1	glycoprotein B pre
24	88.5	5.8	982	2	T43676	hunchback-related
25	88.5	5.8	1071	2	T18597	hypothetical prote
26	88.5	5.8	1650	2	S53457	dominant autoantig
27	88	5.8	678	2	S12456	virD3 protein - Ag
28	88	5.8	791	2	S67265	hypothetical prote
29	88	5.8	888	2	S64016	probable regulator
30	87.5	5.7	578	2	G84015	maltogenic amylase
31	86.5	5.7	407	2	A85191	probable serine pr
32	86.5	5.7	944	2	D82926	hypothetical prote
33	86.5	5.7	967	2	A30325	membrane alanyl am
34	86	5.6	506	2	AB3411	cysteine-tRNA liga
35	86	5.6	633	2	S48956	hypothetical prote
36	86	5.6	739	2	I56187	transcription fact
37	86	5.6	1309	1	BVBYD9	RAD9 protein - yea
38	86	5.6	2509	2	G01880	fatty-acid synthas
39	85.5	5.6	430	1	AJPMQ2	glutamate-ammonia
40	85.5	5.6	2329	2	S44625	C50C3.6 protein -
41	85	5.6	770	1	S30293	transcription fact
42	85	5.6	1566	2	T20058	hypothetical prote
43	84.5	5.5	389	2	S75454	hypothetical prote
44	84	5.5	325	2	JC2008	actin homolog prot
45	83.5	5.5	283	2	D83948	hypothetical prote

ALIGNMENTS

RESULT 1
A54774
ATP binding cassette transporter ABC1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C;Accession: A54774
R;Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A;Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A;Reference number: A54774; MUID:94375008; PMID:8088782
A;Accession: A54774
A;Molecule type: mRNA
A;Residues: 1-2201 <LUC>
A;Cross-references: GB:X75926; NID:g495256; PIDN:CAA53530.1; PID:g495257
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C;Keywords: ATP; duplication; nucleotide binding; P-loop
F;856-1047/Domain: ATP-binding cassette homology <ABC1>
F;873-880/Region: nucleotide-binding motif A (P-loop)
F;1869-2060/Domain: ATP-binding cassette homology <ABC2>
F;1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match 93.3%; Score 1423; DB 2; Length 2201;
Best Local Similarity 93.0%; Pred. No. 1.1e-107;
Matches 264; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

```
Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           |||||||||||||||||||||||||||| || ||||||||||||||||||||||||||||
Db    1311 FGKYPSLELQPWMYNEQYTFVSNDAPEDMGTQELLNALTKDPGFGTRCMEGNPIPDTPCL 1370

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||:|| :||||:|:||||||||||:|||||||||||||||||||||||||||||||||
Db    1371 AGEEDWTISPVPQSIVDLFQNGNWTMKNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1430

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           | |||||:||||||||||||||||||||||||||||||||||||||:|||||| |
Db    1431 QKTADILQNLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHE 1490

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           |||||||||| ||| ||:||||||:|||||| ||||:|||||||||||||||||||||||
Db    1491 VNDAIKQMKKLLKLTKDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLNVIN 1550

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           |||||||||||||| ||||||||||||||||||||||||||||||
Db    1551 NAILRANLQKGENPSQYGITAFNHPLNLTKQQLSEVALMTTSVD 1594
```

RESULT 2
A59189
ATP-binding cassette transporter - human (fragment)
N;Alternate names: KIAA1062 protein
C;Species: Homo sapiens (man)
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 02-Jun-2000

```
C;Accession: A59189
R;Kikuno, R.; Nagase, T.; Ishikawa, K.; Hirosawa, M.; Miyajima, N.; Tanaka, A.; Kotani,
DNA Res. 6, 197-205, 1999
A;Title: Prediction of the coding sequences of unidentified human genes. XIV. The comple
A;Reference number: Z22961; MUID:99397452; PMID:10470851
A;Accession: A59189
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1529 <KIK>
A;Cross-references: GB:AB028985; NID:g5689460; PIDN:BAA83014.1; PID:d1046841; PID:g56894
A;Experimental source: chromosome 9; clone hj03579; clone lib pBluescriptII SK plus; tis
C;Genetics:
A;Map position: 9
A;Note: KIAA1062
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

  Query Match            17.5%;  Score 267;  DB 2;  Length 1529;
  Best Local Similarity  29.2%;  Pred. No. 2.4e-13;
  Matches   80;  Conservative   35;  Mismatches   73;  Indels   86;  Gaps   12;

Qy      44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
           | : |:|                :| | |:|                || | ||:|| :
Db     665 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 724

Qy      72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
           |: : :  :           | ||:         ||   ||  ||| :  | ||| |:|
Db     725 PRLVREPVR-----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDIT 768

Qy     132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
           | |:|:||: |  :               |||  : |  |     |:     |   ::|
Db     769 GHNVSEYLLFTSDRF-----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 814

Qy     192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                                 :  |   :|::||||:|:: ::|| :|||||||| |
Db     815 ----------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 852

Qy     252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
           : ||: ||||  |||:| |   || :  |  | |
Db     853 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 886


RESULT 3
B54774
ATP binding cassette transporter ABC2 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C;Accession: B54774
R;Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A;Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A;Reference number: A54774; MUID:94375008; PMID:8088782
A;Accession: B54774
A;Molecule type: mRNA
A;Residues: 1-1472 <LUC>
A;Cross-references: GB:X75927; NID:g495258; PIDN:CAA53531.1; PID:g495259
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;44-234/Domain: ATP-binding cassette homology <ABC1>
F;61-68/Region: nucleotide-binding motif A (P-loop)
F;1108-1300/Domain: ATP-binding cassette homology <ABC2>
F;1126-1133/Region: nucleotide-binding motif A (P-loop)

  Query Match            16.9%;  Score 257;  DB 2;  Length 1472;
  Best Local Similarity  24.9%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative   40;  Mismatches   98;  Indels  136;  Gaps   15;

Qy       2 GKYPSLELQPWMYNEQYT--------FVSNDAPE-------DTGTLELLNALTKDPGFGT 46
           |  | | | |  |:  ||        : : :  |       |    :|::      | |
Db     519 GDLPPLVLSPSQYH-NYTQPRGNFIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGA 577

Qy      47 RCM-------------------------------------EGNPI--------------- 54
           |:                                     :| |:
Db     578 TCVLKSPANGSLGPMLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDS 637

Qy      55 ---PD----------TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKI 100
              ||           |  || | ||:|| :|: : :  :                | ||:
Db     638 PVXPDEDSLQAWNMSLPPTAGPETWTSAPSLPRLVHEPVR-----------CTCSAQGT 685

Qy     101 KKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEF 160
                 ||    ||  ||| :  | ||| |:|| |:|:||: |  :
Db     686 GFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----------RLH 730

Qy     161 RYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
           |||  : |  | |   |:                          |   :  :  :  |
Db     731 RYGAITFG--NVQKSIPAS----------------------FGARVPPMVRKIAVRRVA 765

Qy     221 KVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQLS-EVAL 278
           :| :||||:|:: ::|| :|||||||| | : ||: | ||  |||:| |   || :  |
Db     766 QVLYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYXITVTNHPMNKTSASLSLDYLL 825

Qy     279 MTTSV 283
             | |
Db     826 QGTDV 830


RESULT 4
T15200
hypothetical protein F12B6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C;Accession: T15200
R;Pauley, A.; Maggi, L.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid F12B6.
A;Reference number: Z18307
A;Accession: T15200
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1447 <PAU>
A;Cross-references: EMBL:AF003138; NID:g2088708; PID:g2088709; PIDN:AAB54153.1; GSPDB
A;Experimental source: strain Bristol N2; clone F12B6
C;Genetics:
A;Gene: CESP:F12B6.1
A;Map position: 1
A;Introns: 79/2; 114/3; 177/1; 224/3; 331/1; 345/3; 373/2; 417/2; 464/1; 536/1; 659/2
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

  Query Match            14.1%;  Score 215.5;  DB 2;  Length 1447;
  Best Local Similarity  24.2%;  Pred. No. 3.7e-09;
  Matches   86;  Conservative   44;  Mismatches  105;  Indels  121;  Gaps   15;

Qy       5 PSLELQPWMYNEQYTFVSN--DAPEDTGTLELLNALTKDPGFGTRCMEG--NPIPDT--- 57
           | | |:  :      |  |  |  |::   ::|:|:   || | || :   | : ||
Db     605 PPLPLETSIMGNHSDFYVNSWDTAENSTANDILHAMFSSPGTGPRCAKDVPNDLLDTMRR 664

Qy      58 ----------------------------PCQ--AGEEEWT--------TAPVPQTIMDL- 78
                                        ||   || ::|         ||:     |
Db     665 ELMFRNRYGFGRNKPAPGVDKDSVDNEYQCQNIQGEFDYTEDISNATYNAPIYCGCEDFG 724

Qy      79 ----FQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRN 134
               ::  |   |                                : || | : ||||||
Db     725 WNCTLEDWKWNETN----------------------------WLRLNTTDRIFDLTGRN 755

Qy     135 ISDY-LVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVND--------AI 185
           :: : |:  : |:   :          |  ||||||  | :|   ||   |         |
Db     756 LTQFRLITRFAQLANTTA-------PFFLGGFSLGHVNQRA--QSQADIDTSKRGWLETI 806

Qy     186 KQMKKHLKL------------------AKDSSADRFLNSLGRFMTGLDTRNNVKVWFNN 226
           | : : :::                  |:: : :: :| |   :  || | ||||||||
Db     807 KDIAQSMRIINLNTTGIEPATPKVLDPFAQNITLNQVVNDL---LQNLDVRENVKVWFNN 863

Qy     227 KGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLT-KQQLSEVALMTT 281
           | |       |:::||:||       :|   ||   |||:| |  | | : ||  |
Db     864 KIWPGFPIASNILSNALLRQE-DYAIDPEDLGILTMNHPMNKTISQTLDQNALKFT 918
```

```
RESULT 5
T14292
glutamate-ammonia ligase (EC 6.3.1.2) - carrot
N;Alternate names: glutamine synthetase
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 03-Jun-2002
C;Accession: T14292
R;Higashi, K.; Kamada, H.
submitted to the EMBL Data Library, August 1997
A;Description: A cDNA sequence of carrot glutamine synthetase.
A;Reference number: Z17963
A;Accession: T14292
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-432 <HIG>
A;Cross-references: EMBL:AF019561; NID:g2454632; PID:g2454633
A;Experimental source: strain US-Harumakigosun; leaf
C;Genetics:
A;Gene: GS2
C;Function:
A;Description: catalyzes the formation of glutamine from ammonia and glutamic acid in th
C;Superfamily: glutamate-ammonia ligase
C;Keywords: ligase

  Query Match             7.1%;  Score 108.5;  DB 2;  Length 432;
  Best Local Similarity   24.5%;  Pred. No. 0.42;
  Matches   71;  Conservative   26;  Mismatches  110;  Indels   83;  Gaps   15;

Qy       3 KYPSLELQPWMYNEQYTFVSNDAPEDTGTLELL-NALTKDPGFG---------TRCMEGN 52
           ::|| ||  | |:   |     || |   : |   |: ||| |         |   :|
Db     105 EHPS-ELPKWNYDGSST---GQAPGDDSEVILYPQAIFKDPFRGGNNILVICDTYTPQGE 160

Qy      53 PIPDTPCQ-----------AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIK 101
           |||                ||  |       |:|   |: ||
Db     161 PIPTNKRHKAAQIFSDAKVLGEVPWFGIEQEYTLMQ--QDVNW----------------- 201

Qy     102 KMLPVCPPG--AGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNE 159
             | |   || | ||     |       ||:|||   |  :
Db     202 ------PLGWNVGGYPGPQGPYYCAAGADKSFGRDISDAHYKACL-------------- 240

Qy     160 FRYGGFSLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDT 216
             | | ::  :| : :|   |  |  ::  :   |:  |:    :|     |  :| ||
Db     241 --YAGINISGTNGEVMPGQWEFQVGPSVGIEAGDHIWCAR-YLLERITEQAGVVLT-LDP 296

Qy     217 RNNVKVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
           :    | |  | |   |              ||  |||  :|:  |: | ||
Db     297 KPIDGDW-NGAGCHTNYSTKSMREEGGFEVIKKAILNLSLRHKEHISAYG 345


RESULT 6
AJBHQ
glutamate-ammonia ligase (EC 6.3.1.2) 2 precursor, chloroplast - barley
N;Alternate names: glutamine synthetase 2
C;Species: Hordeum vulgare (barley)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 03-Jun-2002
C;Accession: S11865; S12687; S14893; A30468; A32363; S05971
R;Stroman, P.; Baima, S.; Casadoro, G.
Plant Mol. Biol. 15, 161-163, 1990
A;Title: A cDNA sequence coding for glutamine synthetase in Hordeum vulgare L.
A;Reference number: S11865; MUID:91355850; PMID:1983297
A;Accession: S11865
A;Molecule type: mRNA
A;Residues: 1-434 <STR>
A;Cross-references: EMBL:X53580
R;Casadoro, G.
submitted to the EMBL Data Library, June 1990
A;Reference number: S12687
A;Accession: S12687
A;Molecule type: mRNA
A;Residues: 1-185,'P',187-434 <CAS>
A;Cross-references: EMBL:X53580; NID:g18987; PIDN:CAA37643.1; PID:g18988
R;Freeman, J.; Marquez, A.J.; Wallsgrove, R.M.; Saarelainen, R.; Forde, B.G.
Plant Mol. Biol. 14, 297-311, 1990
A;Title: Molecular analysis of barley mutants deficient in chloroplast glutamine synt
A;Reference number: S14893; MUID:91346618; PMID:1983286
A;Accession: S14893
A;Molecule type: mRNA
A;Residues: 'AQAVVQAMQCQV',21,'VR',24,'RTA',28-40,'R',42-185,'P',187-434 <FRE>
A;Cross-references: EMBL:X16000; NID:g18985; PIDN:CAA34131.1; PID:g755762
A;Accession: A30468
A;Molecule type: protein
A;Residues: 'X',55-68 <FRE2>
A;Note: a large fraction of the protein has a blocked amino end. In the small fractio
R;Baima, S.; Haegi, A.; Stroman, P.; Casadoro, G.
Carlsberg Res. Commun. 54, 1-9, 1989
A;Title: Characterization of a cDNA clone for barley leaf glutamine synthetase.
A;Reference number: A32363; MUID:89322552; PMID:2473765
A;Accession: A32363
A;Molecule type: mRNA
A;Residues: 48-434 <BAI>
C;Genetics:
A;Map position: 2
C;Superfamily: glutamate-ammonia ligase
C;Keywords: chloroplast; ligase
F;1-53/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;54-434/Product: glutamate-ammonia ligase #status experimental <MAT>

  Query Match             6.6%;  Score 100.5;  DB 1;  Length 434;
  Best Local Similarity   23.1%;  Pred. No. 1.9;
  Matches   66;  Conservative   29;  Mismatches  112;  Indels   79;  Gaps   13;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
           || ||  | |:   |     || ||:  :    |: |||  |         |   :| ||
Db     109 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVICDTYTPQGEPI 164

Qy      55 P-----------DTPCQAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
           |           |    : :|       |:|   :: || :  |
Db     165 PTNKRHMAAQIFSDPKVTSQVQWFGIEQEYTLMQ--RDVNWPLGWP-------------- 208

Qy     104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISD--YLVKTYVQIIAKSLKNKIWVNEFR 161
                 || | ||     |       ||:|||  |    |  |       ::   ::
Db     209 -------VGGYPGPQGPYYCAVGSDKSFGRDISDAHYKACLYAGIEISGTNGEVMPGQWE 261

Qy     162 YG-GFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
           |  | |:|:     :  |: :                :|     |  :| || :
Db     262 YQVGPSVGIDAGDHIWASRYI-----------------LERITEQAGVVLT-LDPKPIQ 302

Qy     221 KVWFNNKGWHAISSFL--------NVINNAILRANLQKGENPSHYG 258
             | |  | |   | |         :||  |||  :|:   : : ||
Db     303 GDW-NGAGCHTNYSTLSMREDGGFDVIKKAILNLSLRHDLHIAAYG 347


RESULT 7
F90070
Clumping factor B [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90070
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-877 <KUR>
A;Cross-references: GB:BA000018; PID:g13702588; PIDN:BAB43728.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: clfB

  Query Match             6.4%;  Score 98;  DB 2;  Length 877;
  Best Local Similarity   21.7%;  Pred. No. 7.8;
  Matches   61;  Conservative   40;  Mismatches  116;  Indels   64;  Gaps   12;
```

```
Qy      15 NEQYTFVSNDAPEDTGTLEL--------LNAL-TKDPGFGTRCMEGNPIPDTPCQAGEEE 65
Db     152 NDANSIATNSELKNSQTLDLPQSSPQTISNAQGTSKPSVRTRAVRSLAVAEPVVNAADAK 211

Qy      66 WTTA---------PVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKK------MLPVCPPG 110
Db     212 GTNVNDKVTASNFKLEKTTFDPNQSGNTFM----AANFTVTDKVKSGDYFTAKLPDSLTG 267

Qy     111 AGGLPPPQRKQNT---ADILQ---DLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGG 164
Db     268 NGDV-DYSNSNNTMPIADIKSTNGDVVAKATYDILTKTYTFVFTDYVNNKENIN----GQ 322

Qy     165 FSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWF 224
Db     323 FSLPLFTDRAKAPKSGTYDA------NINIADEMFNNKITYNYSSPIAGIDKPNGAN--- 373

Qy     225 NNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHP 265
Db     374 --------------ISSQIIGVDTASGQNT--YKQTVFVNP 398
```

RESULT 8
T16057
hypothetical protein F13D11.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
C;Accession: T16057
R;Fulton, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F13D11.
A;Reference number: S69020
A;Accession: T16057
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-908 <FUL>
A;Cross-references: EMBL:U40939; NID:g1072175; PID:g1072177; PIDN:AAA81701.1; CESP:F13D1
C;Genetics:
A;Gene: CESP:F13D11.2
A;Introns: 22/3; 168/3; 539/3; 729/1; 769/3; 839/3; 858/2

Query Match 6.4%; Score 97; DB 2; Length 908;
Best Local Similarity 21.1%; Pred. No. 9.8;
Matches 58; Conservative 32; Mismatches 99; Indels 86; Gaps 10;

```
Qy      53 PIPDTPCQAGEEEWTTAP-------------------------VPQTIMDLFQNGNWT 85
Db     134 PIPITPTQSANVERSNSPSIEEALLLTLSQQQFAERSGTSAFLNIEPKEMSMSSANNNNE 193

Qy      86 MQNPSPACQCSSDKIKKMLPVC-PPGAG--GLPPPQRKQNTADILQDLTG---------- 132
Db     194 EAPASTVSACSTPTTTTSASFCRPPGLGPVALPPTQNGQ-TPMLVCPICGFMCPSKFHFN 252

Qy     133 ----------RNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVN 182
Db     253 SHMNTHGDHQCSMCDYTSRTEGR-LKKHMRESHTVEEQLRAGFESEPAKESASSP----- 306

Qy     183 DAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNA 242
Db     307 -------KNLSLSKDGSATSPINEIFNLST-----------------TMASILDSTNNA 341

Qy     243 ILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVA 277
Db     342 V--SSTSTTEQPS--ALSALTLDMSSTPSLLSTLA 372
```

RESULT 9
AJFBQD
glutamate-ammonia ligase (EC 6.3.1.2) delta precursor, chloroplast - kidney bean
N;Alternate names: glutamine synthetase delta; plasmid-located glutamine synthetase
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 03-Jun-2002
C;Accession: S04031; S24245
R;Lightfoot, D.A.; Green, N.K.; Cullimore, J.V.
Plant Mol. Biol. 11, 191-202, 1988
A;Title: The chloroplast-located glutamine synthetase of Phaseolus vulgaris L.: nucle
A;Reference number: S04031
A;Accession: S04031
A;Molecule type: mRNA
A;Residues: 1-429 <LIG>
A;Cross-references: GB:X12738; NID:g21004; PIDN:CAA31234.1; PID:g21005
R;Cock, J.M.
submitted to the EMBL Data Library, July 1991
A;Reference number: S24245
A;Accession: S24245
A;Molecule type: DNA
A;Residues: 396-429 <COC>
A;Cross-references: EMBL:X61292; NID:g20998; PIDN:CAA43591.1; PID:g805011
C;Superfamily: glutamate-ammonia ligase
C;Keywords: chloroplast; ligase
F;1-57/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;58-429/Product: glutamate-ammonia ligase delta #status predicted <MAT>

Query Match 6.3%; Score 95.5; DB 1; Length 429;
Best Local Similarity 24.1%; Pred. No. 4.8;
Matches 69; Conservative 26; Mismatches 116; Indels 75; Gaps 14;

```
Qy       3 KYPSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFGTRCM---------EGN 52
Db     102 EHPS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVICDAYTPAGE 157

Qy      53 PIPDTPCQAGEEEWTT----APVP-----QTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
Db     158 PIPTNKRHRAAEVFSNPRVIAEVPWFGIEQEYTLLQTNVNWPLGWP-------------- 203

Qy     104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISD--YLVKTYVQIIAKSLKNKIWVNEFR 161
Db     204 -------VGGYPGPQGPYYCSAGADKSFGRDISDAHYKACLFAGINISGTNGEVMPGQWE 256

Qy     162 YG-GFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
Db     257 YQVGPSVGI----------EAGDHIWASRYIL--------ERITEQAG-VVLSLDPKPIE 297

Qy     221 KVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
Db     298 GDW-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHKEHISAYG 342
```

RESULT 10
C64339
hypothetical protein MJ0314 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: C64339
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: C64339
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-263 <BUL>
A;Cross-references: GB:U67486; GB:L77117; NID:g1591031; PIDN:AAB98310.1; PID:g1591034
C;Genetics:
A;Map position: FOR293581-294372
A;Start codon: TTG

Query Match 6.2%; Score 94.5; DB 2; Length 263;
Best Local Similarity 24.6%; Pred. No. 3.1;
Matches 49; Conservative 32; Mismatches 61; Indels 57; Gaps 11;

```
Qy      72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
```

```
Db      43 PQEIIKLYQNG-YTTTEIAKIMKCSHETIRRIL---------------RNNNIDI----- 81

Qy     132 GRNISDYLVKTYVQIIAKSLKN--KIWVNEFRYGGFSLGVSNTQALPPSQEVN------- 182
Db      82 -RKSSESLI----------IKNTKKINLNPSESLAYILGVLNGDGSVNKQESNYVIELKV 130

Qy     183 ---DAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKG---WHA---IS 233
Db     131 TDKDFIEEFKRNL----ENIGFKYINEYVRKFENKKDQYVVRV--RSKGFYYWYKSLNVD 184

Qy     234 SFLNVI-NNAILRANLQKG 251
Db     185 YYMNVIGNNEKLMISWLKG 203


RESULT 11
S18600
glutamate-ammonia ligase (EC 6.3.1.2) precursor, chloroplast (clone lambdaAtgs11) - Arab
N;Alternate names: glutamine synthetase
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 22-Nov-1993 #sequence_revision 12-May-1995 #text_change 03-Jun-2002
C;Accession: S18600
R;Peterman, T.K.; Goodman, H.M.
Mol. Gen. Genet. 230, 145-154, 1991
A;Title: The glutamine synthetase gene family of Arabidopsis thaliana: light-regulation
A;Reference number: S18600; MUID:92079889; PMID:1684022
A;Accession: S18600
A;Molecule type: mRNA
A;Residues: 1-430 <PET>
A;Cross-references: EMBL:S69727; NID:g240069; PIDN:AAB20558.1; PID:g240070
A;Experimental source: clone lambdaAtgs11
C;Genetics:
A;Genome: nuclear
C;Superfamily: glutamate-ammonia ligase
C;Keywords: chloroplast; ligase
F;1-51/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;52-430/Product: glutamate-ammonia ligase #status predicted <MAT>

  Query Match             6.2%;  Score 94.5;  DB 2;  Length 430;
  Best Local Similarity   23.1%;  Pred. No. 5.9;
  Matches   67;  Conservative   25;  Mismatches  111;  Indels   87;  Gaps   14;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
Db     105 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFRDPFRGGNNILVICDTWTPAGEPI 160

Qy      55 P-----------DTPCQAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
Db     161 PTNKRAKAAEIFSNKKVSGEVPWFGIEQEYTLLQ--QNVKWPLGWP-------------- 204

Qy     104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYG 163
Db     205 -------VGAFPGPQGPYYCGVGADKIWGRDISDAHYKACL----------------YA 240

Qy     164 GFSLGVSNTQALPPSQEVN-------DAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDT 216
Db     241 GINISGTNGEVMPGQWEFQVGPSVGIDA----GDHVWCAR-YLLERITEQAGVVLT-LDP 294

Qy     217 RNNVKVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
Db     295 KPIEGDW-NGAGCHTNYSTKSMREEGGFEVIKKAILNLSLRHKEHISAYG 343


RESULT 12
T42737
gp330 protein precursor - rat
N;Alternate names: megalin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C;Accession: T42737
R;Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A;Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of t
A;Reference number: A58173; MUID:95024033; PMID:7937880
A;Accession: T42737
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4660 <SAI>
A;Cross-references: EMBL:L34049; NID:g561852; PID:g561853; PIDN:AAA51369.1
A;Experimental source: strain Sprague-Dawley; kidney
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-bind
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-4660/Product: gp330 protein #status predicted <MAT>

  Query Match             6.2%;  Score 94.5;  DB 2;  Length 4660;
  Best Local Similarity   23.6%;  Pred. No. 1.4e+02;
  Matches   70;  Conservative   37;  Mismatches  111;  Indels   79;  Gaps   19;

Qy       1 FGKYPS---LELQPWM----------YNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTR 47
Db    4302 FGKENKEKVLVVNPWLTQVRIFHQLRYNQS---VSNPCKQVCSHLCLL----RPGGYSCA 4354

Qy      48 CMEGNPI---PDTPCQAGEEEWTTAPVPQTIMDLFQNGN-WTMQNPSPACQCSSDKIKKM 103
Db    4355 CPQGSDFVTGSTVQCDAASELPVTMPPPCRCM---HGGNCYFDENELPKCKCSSGYSGE- 4410

Qy     104 LPVCPPG-AGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRY 162
Db    4411 --YCEVGLSRGIPP--------------GTTMA-VLLTFVIVIIVGAL---VLVGLFHY 4449

Qy     163 GGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLG-RFMTGLDTRNNVK 221
Db    4450 -------RKTGSLLPT------LPKLPSLSSLAKPSE-----NGNGVTFRSGADV--NMD 4489

Qy     222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAL 278
Db    4490 IGVSPFGPETIIDRSMAMNEHFV---MEVGKQP-----VIFENPMYAAKDNTSKVAL 4538


RESULT 13
AJRZQD
glutamate-ammonia ligase (EC 6.3.1.2) delta precursor, chloroplast - rice
N;Alternate names: glutamine synthetase delta
C;Species: Oryza sativa (rice)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 03-Jun-2002
C;Accession: S07471
R;Sakamoto, A.; Ogawa, M.; Masumura, T.; Shibata, D.; Takeba, G.; Tanaka, K.; Fujii,
Plant Mol. Biol. 13, 611-614, 1989
A;Title: Three cDNA sequences coding for glutamine synthetase polypeptides in Oryza s
A;Reference number: S07469; MUID:91370845; PMID:2577497
A;Accession: S07471
A;Molecule type: mRNA
A;Residues: 1-428 <SAK>
A;Cross-references: GB:X14246; NID:g20369; PIDN:CAA32462.1; PID:g20370
C;Superfamily: glutamate-ammonia ligase
C;Keywords: chloroplast; ligase

  Query Match             6.1%;  Score 92.5;  DB 1;  Length 428;
  Best Local Similarity   23.6%;  Pred. No. 8.5;
  Matches   67;  Conservative   30;  Mismatches  112;  Indels   75;  Gaps   14;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
Db     103 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVMCDTYTPAGEPI 158

Qy      55 P----DTPCQAGEEEWTTAPVP-----QTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLP 105
Db     159 PTNKRNRAAQVFSDPKVVSQVPWFGIEQEYTLLQRDVNWPLGWP---------------- 202

Qy     106 VCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGF 165
Db     203 -----VGGYPGPQGPYYCAVGSDKSFGRDISDAHYKACL----------------YAGI 240

Qy     166 SLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKV 222
Db     241 NISGTNGEVMPGQWEYQVGPSVGIEAGDHIWISR-YILERITEQAGVVLT-LDPKPIQGD 298
```

```
Qy    223 WFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
          | |  | |   |           ||  |||  :|:   : | ||
Db    299 W-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHDLHISAYG 341


RESULT 14
B83793
hypothetical protein BH1146 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B83793
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: B83793
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-459 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BAB04865.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1146

  Query Match             6.0%;  Score 91;  DB 2;  Length 459;
  Best Local Similarity   33.3%;  Pred. No. 12;
  Matches   32;  Conservative   11;  Mismatches   29;  Indels   24;  Gaps    6;

Qy    180 EVNDAIKQMKK---HLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWF---NNKGWHAIS 233
          | ::::|::||    | || | | | |:  :| |  |||   : :||    |:|   ||
Db    366 EKHESVKRLKKAALQLSLAIDESLDGFIGEIG-FDLGLD--ESGRVWLLEANSKPGRAI- 421

Qy    234 SFLNVINNAILRANLQKGEN-----PSHYGITAFNH 264
                  |||| |      | |||     |
Db    422 ---------FRHPNLQKEERETLRLPFHYGTFLAKH 448


RESULT 15
S32228
glutamate-ammonia ligase (EC 6.3.1.2) precursor - rape
C;Species: Brassica napus (rape)
C;Date: 07-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 03-Jun-2002
C;Accession: S32228; S32686
R;Ochs, G.; Schock, G.; Wild, A.
submitted to the EMBL Data Library, March 1993
A;Description: Nucleotide sequence of a cDNA encoding chloroplastic glutamine synthetase
A;Reference number: S32228
A;Accession: S32228
A;Molecule type: mRNA
A;Residues: 1-428 <OCH>
A;Cross-references: EMBL:X72751; NID:g296222; PIDN:CAA51280.1; PID:g296223
C;Superfamily: glutamate-ammonia ligase
C;Keywords: ligase

  Query Match             5.9%;  Score 90.5;  DB 2;  Length 428;
  Best Local Similarity   23.9%;  Pred. No. 12;
  Matches   68;  Conservative   28;  Mismatches  113;  Indels   75;  Gaps   14;

Qy      5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
          || ||  | |:   |     || ||:  :    |: :||  |         |    | ||
Db    103 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFRDPFRGGNNILVICDTYTPAGEPI 158

Qy     55 P-DTPCQAGE--------EEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLP 105
          | :   :| |        ||    : |    |  | || :  |
Db    159 PTNKRARAAEIFSNKKVNEEIPWFGIEQEYTLLQPNVNWPLGWP---------------- 202

Qy    106 VCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGF 165
                |  | ||         :   ||:|||   |  :                | |
Db    203 -----VGAYPGPQGPYYCGVGAEKSWGRDISDAHYKACL----------------YAGI 240

Qy    166 SLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKV 222
          ::  :| : :|   |  |  ::  :   |:  |:    :|     |  :| || :
```

```
Db    241 NISGTNGEVMPGQWEFQVGPSVGIEAGDHVWCAR-YLLERITEQAGVVLT-LDPKPIEGD 298

Qy    223 WFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
          | |  | |   |           ||  |||  :|:   |: | ||
Db    299 W-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHMEHISAYG 341


Search completed: February  4, 2003, 09:40:38
Job time : 25 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:38:24 ; Search time 11 Seconds
(without alignments)
1070.843 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1513	99.2	2261	1	ABC1_HUMAN	O95477 homo sapien
2	1423	93.3	2261	1	ABC1_MOUSE	P41233 mus musculu
3	733.5	48.1	2273	1	ABCR_HUMAN	P78363 homo sapien
4	267	17.5	2436	1	ABC2_HUMAN	Q9bzc7 homo sapien
5	264	17.3	2434	1	ABC2_MOUSE	P41234 mus musculu
6	108.5	7.1	432	1	GLN2_DAUCA	O22506 daucus caro
7	98.5	6.5	434	1	GLN2_HORVU	P13564 hordeum vul
8	98.5	6.5	2083	1	DYSF_MOUSE	Q9esd7 mus musculu
9	95.5	6.3	429	1	GLN4_PHAVU	P15102 phaseolus v
10	94.5	6.2	263	1	Y314_METJA	Q57762 methanococc
11	94.5	6.2	430	1	GLN2_ARATH	Q43127 arabidopsis
12	94.5	6.2	2080	1	DYSF_HUMAN	O75923 homo sapien
13	94.5	6.2	4660	1	LRP2_RAT	P98158 rattus norv
14	92.5	6.1	428	1	GLN2_ORYSA	P14655 oryza sativ
15	90.5	5.9	428	1	GLNC_BRANA	Q42624 brassica na
16	89.5	5.9	423	1	GLNC_MAIZE	P25462 zea mays (m
17	89.5	5.9	596	1	PPOB_LYCES	Q08304 lycopersico
18	89	5.8	903	1	VGLB_HSV1F	P06436 herpes simp
19	89	5.8	1597	1	GTF1_STRDO	P11001 streptococc
20	88.5	5.8	363	1	YK57_YEAST	P36157 saccharomyc
21	88.5	5.8	982	1	HBL1_CAEEL	Q9xyd3 caenorhabdi
22	88	5.8	678	1	VID3_AGRRH	P13463 agrobacteri
23	88	5.8	888	1	YGB4_YEAST	P25339 saccharomyc
24	87.5	5.7	428	1	GLN2_MEDSA	Q9xq94 medicago sa
25	87.5	5.7	3148	1	HD_FUGRU	P51112 fugu rubrip
26	87	5.7	2768	1	THYG_HUMAN	P01266 homo sapien
27	86.5	5.7	944	1	Y166_UREPA	Q9pqx7 ureaplasma
28	86	5.6	633	1	BZZ1_YEAST	P38822 saccharomyc
29	86	5.6	1309	1	RAD9_YEAST	P14737 saccharomyc
30	85.5	5.6	430	1	GLN2_PEA	P08281 pisum sativ
31	85.5	5.6	2212	1	RRPL_EBOZM	Q05318 ebola virus
32	85.5	5.6	2329	1	YLJ6_CAEEL	P34369 caenorhabdi
33	85	5.6	770	1	OCT1_MOUSE	P25425 mus musculu
34	84	5.5	322	1	ACT_PROCL	P45521 procambarus
35	83.5	5.5	353	1	VM17_BORHE	P32777 borrelia he
36	83.5	5.5	356	1	GLN3_ORYSA	P14656 oryza sativ
37	83.5	5.5	966	1	AMPN_HUMAN	P15144 homo sapien
38	83.5	5.5	1377	1	CID_DROME	P19538 drosophila
39	83.5	5.5	3224	1	RBP2_HUMAN	P49792 homo sapien
40	83	5.4	559	1	3BP2_MOUSE	Q06649 mus musculu
41	83	5.4	795	1	SYFB_BUCAI	P57230 buchnera ap
42	83	5.4	913	1	VGLB_PRVIF	P08355 pseudorabie
43	82.5	5.4	355	1	GLN4_MAIZE	P38562 zea mays (m
44	82.5	5.4	356	1	GLN3_MAIZE	P38561 zea mays (m
45	82.5	5.4	522	1	IMA_DROME	P52295 drosophila

ALIGNMENTS

RESULT 1
ABC1_HUMAN
ID ABC1_HUMAN STANDARD; PRT; 2261 AA.
AC O95477; Q9UN08; Q9UN07; Q9UN06; Q9NQV4; Q9UN09; Q96T85; Q96S56;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette
DE transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux
DE regulatory protein).
GN ABCA1 OR ABC1 OR CERP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345099; PubMed=10884428;
RA Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y.,
RA Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P.,
RA Haudenschild C.C., Prades C., Chimini G., Blackmon E.E.,
RA Francois T.L., Duverger N., Rubin E.M., Rosier M., Denefle P.,
RA Fredrickson D.S., Brewer H.B. Jr.;
RT "Complete genomic sequence of the human ABCA1 gene: analysis of the
RT human and mouse ATP-binding cassette A promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Schwartz K., Lawn R.M., Wade D.P.;
RT "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are
RT regulated by LXR.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21251004; PubMed=11352567;
RA Qiu Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis
RT studies revealing novel regulatory sequences.";
RL Genomics 73:66-76(2001).
RN [4]
RP SEQUENCE FROM N.A.
RA Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadanami K., Kidera A.,
RA Kioka N., Amachi T., Yokoyama S., Ueda K.;
RT "A new topological model of functional human ABCA1-signal peptide
RT cleavage and glycosylation of a large extracellular domain.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE=99194549; PubMed=10092505;
RA Langmann T., Klucken J., Reil M., Liebisch G., Luciani M.F.,
RA Chimini G., Kaminski W.E., Schmitz G.;
RT "Molecular cloning of the human ATP-binding cassette transporter 1
RT (hABC1): evidence for sterol-dependent regulation in macrophages.";
RL Biochem. Biophys. Res. Commun. 257:29-33(1999).
RN [6]

RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE=99364413; PubMed=10431238;
RA Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
RA Deleuze J.-F., Brewer H.B., Duverger N., Denefle P., Assmann G.;
RT "Tangier disease is caused by mutations in the gene encoding
RT ATP-binding cassette transporter 1.";
RL Nat. Genet. 22:352-355(1999).
RN [7]
RP VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
RX MEDLINE=20001430; PubMed=10533863;
RA Marcil M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
RA Ouelette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,
RA Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;
RT "Mutations in the ABC1 gene in familial HDL deficiency with defective
RT cholesterol efflux.";
RL Lancet 354:1341-1346(1999).
RN [8]
RP VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
RX MEDLINE=99364411; PubMed=10431236;
RA Brooks-Wilson A., Marcil M., Clee S.M., Zhang L.-H., Roomp K.,
RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
RA Loubser O., Ouelette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
RA Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,
RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Mutations in ABC1 in Tangier disease and familial high-density
RT lipoprotein deficiency.";
RL Nat. Genet. 22:336-345(1999).
RN [9]
RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
RP MET-883.
RX MEDLINE=99364412; PubMed=10431237;
RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
RA Diederich W., Drobnik W., Barlage S., Buechler C.,
RA Porsch-Oezcueruemez M., Kaminski W.E., Hahmann H.W., Oette K.,
RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;
RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
RT Tangier disease.";
RL Nat. Genet. 22:347-351(1999).
RN [10]
RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU-693
RP DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
RX MEDLINE=20540002; PubMed=11086027;
RA Clee S.M., Kastelein J.J.P., van Dam M., Marcil M., Roomp K.,
RA Zwarts K.Y., Collins J.A., Roelants R., Tamasawa N., Stulc T.,
RA Suda T., Ceska R., Boucher B., Rondeau C., DeSouich C.,
RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
RA Hayden M.R.;
RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
RT coronary artery disease in ABCA1 heterozygotes.";
RL J. Clin. Invest. 106:1263-1270(2000).
RN [11]
RP VARIANTS TD ASN-1289 AND HIS-1800.
RX MEDLINE=20171564; PubMed=10706591;
RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
RA Van Eerdewegh P., Goldkamp A.L., Thurston L.M., FitzGerald M.G.,
RA Yasek-McKenna D., O'Neill G., Eberhart G.P., Weiffenbach B.,
RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
RT tangier disease kindreds.";
RL J. Lipid Res. 41:433-441(2000).
RN [12]
RP VARIANT TD ASP-1046, VARIANT FHA CYS-230, AND VARIANTS LYS-219;
RP ILE-825; MET-883 AND LYS-1587.
RX MEDLINE=20396633; PubMed=10938021;
RA Wang J., Burnett J.R., Near S., Young K., Zinman B., Hanley A.J.G.,
RA Connelly P.W., Harris S.B., Hegele R.A.;
RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";
RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).
RN [13]
RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
RX MEDLINE=21157002; PubMed=11257260;
RA Bertolini S., Pisciotta L., Seri M., Cusano R., Cantafora A.,
RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.;
RT "A point mutation in ABC1 gene in a patient with severe premature
RT coronary heart disease and mild clinical phenotype of Tangier
RT disease.";
RL Atherosclerosis 154:599-605(2001).
RN [14]
RP VARIANTS LYS-219; MET-883 AND ASP-1172.
RX MEDLINE=21157003; PubMed=11257261;
RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,
RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
RA Schmitz G.;
RT "Common variants in the gene encoding ATP-binding cassette transporter
RT 1 in men with low HDL cholesterol levels and coronary heart disease.";
RL Atherosclerosis 154:607-611(2001).
RN [15]
RP VARIANT TD LEU-1506.
RX MEDLINE=21369429; PubMed=11476961;
RA Lapicka-Bodzioch K., Bodzioch M., Kruell M., Kielar D., Probst M.,
RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
RA Suttorp N., Schmitz G.;
RT "Homogeneous assay based on 52 primer sets to scan for mutations of
RT the ABCA1 gene and its application in genetic analysis of a new
RT patient with familial high-density lipoprotein deficiency syndrome.";
RL Biochim. Biophys. Acta 1537:42-48(2001).
RN [16]
RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS-219.
RX MEDLINE=21369433; PubMed=11476965;
RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
RA Mawatari K., Imamura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;
RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
RT disease and familial high density lipoprotein deficiency with
RT coronary heart disease.";
RL Biochim. Biophys. Acta 1537:71-78(2001).
RN [17]
RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
RP MET-883; ASP-1172; LYS-1587 AND CYS-1731.
RX MEDLINE=21138379; PubMed=11238261;
RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Common genetic variation in ABCA1 is associated with altered
RT lipoprotein levels and a modified risk for coronary artery disease.";
RL Circulation 103:1198-1205(2001).
RN [18]
RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611.
RX MEDLINE=21645894; PubMed=11785958;
RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
RA Wellington C., Yoshida Y., Misugi S., Hayden M.R., Egashira T.,
RA Yamashita S., Matsuzawa Y.;
RT "Expression and functional analyses of novel mutations of ATP-binding
RT cassette transporter-1 in japanese patients with high-density
RT lipoprotein deficiency.";
RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
CC -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
CC MACROPHAGES.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
CC DEFICIENCY TYPE I (HDLD1), ALSO KNOWN AS TANGIER DISEASE (TD). TD
CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
CC DISEASE (CAD).
CC -!- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein

```
  Query Match             99.2%;  Score 1513;  DB 1;  Length 2261;
  Best Local Similarity   99.3%;  Pred. No. 7.5e-120;
  Matches  282;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1431 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1491 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1551 VNDATKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           ||||||||||||||||||||||||||||||||||| ||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAPMTTSVD 1654


RESULT 2
ABC1_MOUSE
ID   ABC1_MOUSE     STANDARD;      PRT;  2261 AA.
AC   P41233;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette
DE   transporter 1) (ATP-binding cassette 1) (ABC-1).
GN   ABCA1 OR ABC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2; TISSUE=Macrophage;
RX   MEDLINE=94375008; PubMed=8088782;
RA   Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;
RT   "Cloning of two novel ABC transporters mapping on human chromosome
RT   9.";
RL   Genomics 21:150-159(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;
RT   "Human and mouse ABCA1 comparative sequencing and transgenesis studies
RT   identify potential regulatory sequences.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
CC       TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC       TRANSPORT (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST
CC       LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
CC   -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC       EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC       ATP BINDING CASSETTE (ABC) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75926; CAA53530.1; ALT_INIT.
DR   EMBL; AF287263; AAG39073.1; ALT_INIT.
DR   MGD; MGI:99607; Abca1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Glycoprotein; Transmembrane; Transport.
FT   TRANSMEM     26     42       POTENTIAL.
FT   TRANSMEM    640    656       POTENTIAL.
FT   TRANSMEM    690    706       POTENTIAL.
FT   TRANSMEM    717    733       POTENTIAL.
FT   TRANSMEM    749    765       POTENTIAL.
FT   TRANSMEM    771    787       POTENTIAL.
FT   TRANSMEM   1041   1057       POTENTIAL.
FT   TRANSMEM   1351   1367       POTENTIAL.
FT   TRANSMEM   1661   1677       POTENTIAL.
FT   TRANSMEM   1708   1724       POTENTIAL.
FT   TRANSMEM   1737   1753       POTENTIAL.
FT   TRANSMEM   1775   1791       POTENTIAL.
FT   TRANSMEM   1854   1870       POTENTIAL.
FT   NP_BIND     933    940       ATP (POTENTIAL).
FT   NP_BIND    1946   1953       ATP (POTENTIAL).
FT   CARBOHYD     14     14       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1144   1144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1294   1294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1453   1453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1499   1499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1504   1504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1637   1637       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2044   2044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2238   2238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1567   1568       MISSING (IN REF. 2).
FT   CONFLICT   2024   2024       MISSING (IN REF. 2).
SQ   SEQUENCE   2261 AA;  254011 MW;  FAE62B21FD1D09F9 CRC64;

  Query Match             93.3%;  Score 1423;  DB 1;  Length 2261;
  Best Local Similarity   93.0%;  Pred. No. 3.2e-112;
  Matches  264;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           |||||||||||||||||||||||||| || ||||||||||||||||||||||||||||
Db    1371 FGKYPSLELQPWMYNEQYTFVSNDAPEDMGTQELLNALTKDPGFGTRCMEGNPIPDTPCL 1430

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
           ||||:|| :||||:|:|||||||||||:|||||||||||||||||||||||||||||||
Db    1431 AGEEDWTISPVPQSIVDLFQNGNWTMKNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 1490

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           | |||||:||||||||||||||||||||||||||||||||||||||||:|||||| |
Db    1491 QKTADILQNLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNSQALPPSHE 1550

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           |||||||||| ||| ||:||||||:|||||| ||||:|||||||||||||||||||||||
Db    1551 VNDAIKQMKKLLKLTKDTSADRFLSSLGRFMAGLDTKNNVKVWFNNKGWHAISSFLNVIN 1610

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           ||||||||||||||| |||||||||||||||||||||||||||||
Db    1611 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 1654
```

```
Db   1611 NAILRANLQKGENPSQYGITAFNHPLNLTKQQLSEVALMTTSVD 1654


RESULT 3
ABCR_HUMAN
ID   ABCR_HUMAN     STANDARD;      PRT;  2273 AA.
AC   P78363; O60438; O60915; O15112;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Retinal-specific ATP-binding cassette transporter (RIM ABC
DE   transporter) (RIM protein) (RMP) (Stargardt disease protein).
GN   ABCA4 OR ABCR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., VARIANTS STGD, AND VARIANTS HIS-846 AND GLN-943.
RX   MEDLINE=97207641; PubMed=9054934;
RA   Allikmets R., Singh N., Sun H., Shroyer N.F., Hutchinson A.,
RA   Chidambaram A., Gerrard B., Baird L., Stauffer D., Peiffer A.,
RA   Rattner A., Smallwood P.M., Li Y., Anderson K.L., Lewis R.A.,
RA   Nathans J., Leppert M., Dean M., Lupski J.R.;
RT   "A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is
RT   mutated in recessive Stargardt macular dystrophy.";
RL   Nat. Genet. 15:236-246(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97345663; PubMed=9202155;
RA   Azarian S.M., Travis G.H.;
RT   "The photoreceptor rim protein is an ABC transporter encoded by the
RT   gene for recessive Stargardt's disease (ABCR).";
RL   FEBS Lett. 409:247-252(1997).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS STGD TRP-18 AND CYS-212.
RX   MEDLINE=98163759; PubMed=9503029;
RA   Gerber S., Rozet J.-M., van de Pol T.J.R., Hoyng C.B., Munnich A.,
RA   Blankenagel A., Kaplan J., Cremers F.P.M.;
RT   "Complete exon-intron structure of the retina-specific ATP binding
RT   transporter gene (ABCR) allows the identification of novel mutations
RT   underlying Stargardt disease.";
RL   Genomics 48:139-142(1998).
RN   [4]
RP   SEQUENCE FROM N.A., AND VARIANTS STGD.
RX   MEDLINE=98141123; PubMed=9490294;
RA   Nasonkin I., Illing M., Koehler M.R., Schmid M., Molday R.S.,
RA   Weber B.H.F.;
RT   "Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1
RT   and identification of novel mutations in Stargardt's disease.";
RL   Hum. Genet. 102:21-26(1998).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=99175213; PubMed=10075733;
RA   Sun H., Molday R.S., Nathans J.;
RT   "Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR,
RT   the photoreceptor-specific ATP-binding cassette transporter
RT   responsible for Stargardt disease.";
RL   J. Biol. Chem. 274:8269-8281(1999).
RN   [6]
RP   DISEASE.
RX   MEDLINE=98133912; PubMed=9466990;
RA   Cremers F.P.M., van de Pol D.J.R., van Driel M.A., den Hollander A.I.,
RA   van Haren F.J.J., Knoers N.V.A.M., Tijmes N., Bergen A.A.B.,
RA   Rohrschneider K., Blankenagel A., Pinckers A.J.L.G., Deutman A.F.,
RA   Hoyng C.B.;
RT   "Autosomal recessive retinitis pigmentosa and cone-rod dystrophy
RT   caused by splice site mutations in the Stargardt's disease gene
RT   ABCR.";
RL   Hum. Mol. Genet. 7:355-362(1998).
RN   [7]
RP   VARIANTS ARMD2, AND VARIANTS.
RX   MEDLINE=97442530; PubMed=9295268;
RA   Allikmets R., Shroyer N.F., Singh N., Seddon J.M., Lewis R.A.,
RA   Bernstein P.S., Peiffer A., Zabriskie N.A., Li Y., Hutchinson A.,
RA   Dean M., Lupski J.R., Leppert M.;
RT   "Mutation of the Stargardt disease gene (ABCR) in age-related macular
RT   degeneration.";
RL   Science 277:1805-1807(1997).
RN   [8]
RP   VARIANTS STGD W-18; C-212; H-636; M-1019; V-1038; C-1108; W-1640;
RP   S-1977 AND H-2107, AND VARIANTS FFM P-11; P-541; V-1038; E-1091;
RP   C-1508; F-1970 AND R-1971.
RX   MEDLINE=98454319; PubMed=9781034;
RA   Rozet J.-M., Gerber S., Souied E., Perrault I., Chatelin S., Ghazi I.,
RA   Leowski C., Dufier J.-L., Munnich A., Kaplan J.;
RT   "Spectrum of ABCR gene mutations in autosomal recessive macular
RT   dystrophies.";
RL   Eur. J. Hum. Genet. 6:291-295(1998).
RN   [9]
RP   VARIANTS STGD.
RX   MEDLINE=99138655; PubMed=9973280;
RA   Lewis R.A., Shroyer N.F., Singh N., Allikmets R., Hutchinson A.,
RA   Li Y., Lupski J.R., Leppert M., Dean M.;
RT   "Genotype/phenotype analysis of a photoreceptor-specific ATP-binding
RT   cassette transporter gene, ABCR, in Stargardt disease.";
RL   Am. J. Hum. Genet. 64:422-434(1999).
RN   [10]
RP   VARIANTS STGD, AND VARIANTS.
RX   MEDLINE=99192348; PubMed=10090887;
RA   Maugeri A., van Driel M.A., van de Pol D.J.R., Klevering B.J.,
RA   van Haren F.J.J., Tijmes N., Bergen A.A.B., Rohrschneider K.,
RA   Blankenagel A., Pinckers A.J.L.G., Dahl N., Brunner H.G.,
RA   Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT   "The 2588G-->C mutation in the ABCR gene is a mild frequent founder
RT   mutation in the western European population and allows the
RT   classification of ABCR Mutations in patients with Stargardt disease.";
RL   Am. J. Hum. Genet. 64:1024-1035(1999).
RN   [11]
RP   VARIANT STGD TYR-54, AND VARIANT ALA-863.
RX   MEDLINE=20077755; PubMed=10612508;
RA   Zhang K., Garibaldi D.C., Kniazeva M., Albini T., Chiang M.F.,
RA   Kerrigan M., Sunness J.S., Han M., Allikmets R.;
RT   "A novel mutation in the ABCR gene in four patients with autosomal
RT   recessive Stargardt disease.";
RL   Am. J. Ophthalmol. 128:720-724(1999).
RN   [12]
RP   VARIANTS STGD V-60; R-206; N-300; P-541; A-849; P-974; V-1038; C-1108;
RP   L-1408; R-1488; D-1652; P-1729; E-1961; W-2038; W-2077; H-2107; R-2128
RP   AND Y-2150.
RX   MEDLINE=99221420; PubMed=10206579;
RA   Fishman G.A., Stone E.M., Grover S., Derlacki D.J., Haines H.L.,
RA   Hockey R.R.;
RT   "Variation of clinical expression in patients with Stargardt dystrophy
RT   and sequence variations in the ABCR gene.";
RL   Arch. Ophthalmol. 117:504-510(1999).
RN   [13]
RP   VARIANTS GLU-1961 AND ASN-2177.
RX   MEDLINE=20349288; PubMed=10880298;
RA   Allikmets R., Tammur J., Hutchinson A., Lewis R.A., Shroyer N.F.,
RA   Dalakishvili K., Lupski J.R., Steiner K., Pauleikhoff D., Holz F.G.,
RA   Weber B.H.F., Dean M., Atkinson A., Gail M.H., Bernstein P.S.,
RA   Singh N., Peiffer A., Zabriskie N.A., Leppert M., Seddon J.M.,
RA   Zhang K., Sunness J.S., Udar N.S., Yelchits S., Silva-Garcia R.,
RA   Small K.W., Simonelli F., Testa F., D'Urso M., Brancato R.,
RA   Rinaldi E., Ingvast S., Taube A., Wadelius C., Souied E., Ducroq D.,
RA   Kaplan J., Assink J.J.M., ten Brink J.B., de Jong P.T.V.M.,
RA   Bergen A.A.B., Maugeri A., van Driel M.A., Hoyng C.B., Cremers F.P.M.,
RA   Paloma E., Coco R., Balcells S., Gonzalez-Duarte R., Kermani S.,
RA   Stanga P., Bhattacharya S.S., Bird A.C.;
RT   "Further evidence for an association of ABCR alleles with age-related
RT   macular degeneration.";
RL   Am. J. Hum. Genet. 67:487-491(2000).
RN   [14]
RP   VARIANTS STGD E-60; T-60; E-65; L-68; R-72; C-212; S-230; S-247;
RP   V-328; K-471; P-541; Q-572; R-607; K-635; C-653; Y-764; R-765; A-901;
```

```
RP   I-959; K-1036; V-1038; P-1063; D-1087; C-1097; C-1108; L-1380; K-1399;
RP   P-1430; V-1440; H-1443; L-1486; Y-1488; M-1537; P-1689; L-1705;
RP   T-1733; R-1748; P-1763; K-1885; H-1898; E-1961; R-1975; S-1977; G-2077
RP   W-2077 AND V-2241, AND VARIANTS Q-152; H-212; R-423; I-552; R-914;
RP   Q-943; T-1562; I-1868; M-1921; L-1948; F-1970; A-2059; N-2177 AND
RP   V-2216.
RX   MEDLINE=20442027; PubMed=10958763;
RA   Rivera A., White K., Stoehr H., Steiner K., Hemmrich N., Grimm T.,
RA   Jurklies B., Lorenz B., Scholl H.P.N., Apfelstedt-Sylla E.,
RA   Weber B.H.F.;
RT   "A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene
RT   in Stargardt disease and age-related macular degeneration.";
RL   Am. J. Hum. Genet. 67:800-813(2000).
RN   [15]
RP   VARIANTS CORD3 GLU-65; CYS-212; PRO-541; ALA-863; GLY-863 DEL;
RP   VAL-1038; LYS-1122; TYR-1490 AND ASP-1598.
RX   MEDLINE=20442040; PubMed=10958761;
RA   Maugeri A., Klevering B.J., Rohrschneider K., Blankenagel A.,
RA   Brunner H.G., Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT   "Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal
RT   recessive cone-rod dystrophy.";
RL   Am. J. Hum. Genet. 67:960-966(2000).
RN   [16]
RP   VARIANTS STGD ASP-340; GLN-572; ALA-863; SER-965; VAL-1038; ALA-1780
RP   AND HIS-1898, AND VARIANT GLN-943.
RX   MEDLINE=20208356; PubMed=10746567;
RA   Shroyer N.F., Lewis R.A., Lupski J.R.;
RT   "Complex inheritance of ABCR mutations in Stargardt disease: linkage
RT   disequilibrium, complex alleles, and pseudodominance.";
RL   Hum. Genet. 106:244-248(2000).
RN   [17]
RP   VARIANTS STGD.
RX   MEDLINE=20098082; PubMed=10634594;
RA   Papaioannou M., Ocaka L., Bessant D., Lois N., Bird A.C., Payne A.,
RA   Bhattacharya S.S.;
RT   "An analysis of ABCR mutations in British patients with recessive
RT   retinal dystrophies.";
RL   Invest. Ophthalmol. Vis. Sci. 41:16-19(2000).
RN   [18]
RP   VARIANTS STGD C-212; D-767; I-897; V-1038; K-1087; K-1399; Q-1640 AND
RP   E-1961, AND VARIANT HIS-212.
RX   MEDLINE=20174852; PubMed=10711710;
RA   Simonelli F., Testa F., de Crecchio G., Rinaldi E., Hutchinson A.,
RA   Atkinson A., Dean M., D'Urso M., Allikmets R.;
RT   "New ABCR mutations and clinical phenotype in Italian patients with
RT   Stargardt disease.";
RL   Invest. Ophthalmol. Vis. Sci. 41:892-897(2000).
RN   [19]
RP   CHARACTERIZATION OF VARIANTS, AND MUTAGENESIS OF GLY-966; LYS-969;
RP   GLY-1975 AND LYS-1978.
RX   MEDLINE=20472331; PubMed=11017087;
RA   Sun H., Smallwood P.M., Nathans J.;
RT   "Biochemical defects in ABCR protein variants associated with human
RT   retinopathies.";
RL   Nat. Genet. 26:242-246(2000).
RN   [20]
RP   VARIANT STGD ASN-972, AND VARIANTS GLN-943; ILE-1868 AND LEU-1948.
RX   MEDLINE=21478761; PubMed=11594993;
RA   Eksandh L., Ekstroem U., Abrahamson M., Bauer B., Andreasson S.;

  Query Match             48.1%;  Score 733.5;  DB 1;  Length 2273;
  Best Local Similarity   48.2%;  Pred. No. 9.1e-54;
  Matches  146;  Conservative   35;  Mismatches   83;  Indels   39;  Gaps    4;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           ||:||:| | ||:| :|||| | | |       | : |   |||| ||::   :|: ||
Db    1397 FGEYPALTLHPWIYGQQYTFFSMDEPGSEQFTVLADVLLNKPGFGNRCLKEGWLPEYPC- 1455

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             | |  |   |  |||   ||  ||||:|:||: :   ||| || ||||||||||
Db    1456 GNSTPWKTPSVSPNITQLFQKQKWTQVNPSPSCRCSTREKLTMLPECPEGAGGLPPPQRT 1515

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           | : :|||||| |||||:|||||  :|  |||:| |||| |||| |:|      ||
Db    1516 QRSTEILQDLTDRNISDFLVKTYPALIRSSLKSKFWVNEQRYGGISIG----GKLPVVPI 1571

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGR------------------FMTGLDTRNNVK 221
             :|:                  ||: |||                  |:  |:| :|:|
Db    1572 TGEALV--------------GFLSDLGRIMNVSGGPITREASKEIPDFLKHLETEDNIK 1616

Qy     222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTT 281
           |||||||||: ||||| :|||||:| |  :|  ||||  : ||||||:||||: ::||
Db    1617 VWFNNKGWHALVSFLNVAHNAILRASLPKDRSPEEYGITVISQPLNLTKEQLSEITVLTT 1676

Qy     282 SVD 284
           |||
Db    1677 SVD 1679


RESULT 4
ABC2_HUMAN
ID   ABC2_HUMAN     STANDARD;      PRT;  2436 AA.
AC   Q9BZC7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette
DE   transporter 2) (ATP-binding cassette 2).
GN   ABCA2 OR ABC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11178988;
RA   Kaminski W.E., Piehler A., Pullmann K., Porsch-Ozcurumez M., Duong C.,
RA   Bared G.M., Buchler C., Schmitz G.;
RT   "Complete coding sequence, promoter region, and genomic structure of
RT   the human ABCA2 gene and evidence for sterol-dependent regulation in
RT   macrophages.";
RL   Biochem. Biophys. Res. Commun. 281:249-258(2001).
CC   -!- FUNCTION: PROBABLE TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN
CC       FOUND YET. MAY HAVE A ROLE IN MACROPHAGE LIPID METABOLISM AND
CC       NEURAL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF327705; AAK14335.1; -.
DR   EMBL; AF327658; AAK14335.1; JOINED.
DR   EMBL; AF327659; AAK14335.1; JOINED.
DR   EMBL; AF327660; AAK14335.1; JOINED.
DR   EMBL; AF327661; AAK14335.1; JOINED.
DR   EMBL; AF327662; AAK14335.1; JOINED.
DR   EMBL; AF327663; AAK14335.1; JOINED.
DR   EMBL; AF327664; AAK14335.1; JOINED.
DR   EMBL; AF327665; AAK14335.1; JOINED.
DR   EMBL; AF327666; AAK14335.1; JOINED.
DR   EMBL; AF327667; AAK14335.1; JOINED.
DR   EMBL; AF327668; AAK14335.1; JOINED.
DR   EMBL; AF327669; AAK14335.1; JOINED.
DR   EMBL; AF327670; AAK14335.1; JOINED.
DR   EMBL; AF327671; AAK14335.1; JOINED.
DR   EMBL; AF327672; AAK14335.1; JOINED.
DR   EMBL; AF327673; AAK14335.1; JOINED.
DR   EMBL; AF327674; AAK14335.1; JOINED.
DR   EMBL; AF327675; AAK14335.1; JOINED.
DR   EMBL; AF327676; AAK14335.1; JOINED.
```

```
DR   EMBL; AF327677; AAK14335.1; JOINED.
DR   EMBL; AF327678; AAK14335.1; JOINED.
DR   EMBL; AF327679; AAK14335.1; JOINED.
DR   EMBL; AF327680; AAK14335.1; JOINED.
DR   EMBL; AF327681; AAK14335.1; JOINED.
DR   EMBL; AF327682; AAK14335.1; JOINED.
DR   EMBL; AF327683; AAK14335.1; JOINED.
DR   EMBL; AF327684; AAK14335.1; JOINED.
DR   EMBL; AF327685; AAK14335.1; JOINED.
DR   EMBL; AF327686; AAK14335.1; JOINED.
DR   EMBL; AF327687; AAK14335.1; JOINED.
DR   EMBL; AF327688; AAK14335.1; JOINED.
DR   EMBL; AF327689; AAK14335.1; JOINED.
DR   EMBL; AF327690; AAK14335.1; JOINED.
DR   EMBL; AF327691; AAK14335.1; JOINED.
DR   EMBL; AF327692; AAK14335.1; JOINED.
DR   EMBL; AF327693; AAK14335.1; JOINED.
DR   EMBL; AF327694; AAK14335.1; JOINED.
DR   EMBL; AF327695; AAK14335.1; JOINED.
DR   EMBL; AF327696; AAK14335.1; JOINED.
DR   EMBL; AF327697; AAK14335.1; JOINED.
DR   EMBL; AF327698; AAK14335.1; JOINED.
DR   EMBL; AF327699; AAK14335.1; JOINED.
DR   EMBL; AF327700; AAK14335.1; JOINED.
DR   EMBL; AF327701; AAK14335.1; JOINED.
DR   EMBL; AF327702; AAK14335.1; JOINED.
DR   EMBL; AF327703; AAK14335.1; JOINED.
DR   EMBL; AF327704; AAK14335.1; JOINED.
DR   Genew; HGNC:32; ABCA2.
DR   MIM; 600047; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein.
FT   TRANSMEM     21     40       POTENTIAL.
FT   TRANSMEM    706    728       POTENTIAL.
FT   TRANSMEM    749    771       POTENTIAL.
FT   TRANSMEM    786    808       POTENTIAL.
FT   TRANSMEM    813    835       POTENTIAL.
FT   TRANSMEM    850    872       POTENTIAL.
FT   TRANSMEM    892    914       POTENTIAL.
FT   TRANSMEM   1793   1815       POTENTIAL.
FT   TRANSMEM   1846   1865       POTENTIAL.
FT   TRANSMEM   1875   1897       POTENTIAL.
FT   TRANSMEM   1904   1926       POTENTIAL.
FT   TRANSMEM   1988   2010       POTENTIAL.
FT   NP_BIND    1025   1032       ATP (POTENTIAL).
FT   NP_BIND    2088   2095       ATP (POTENTIAL).
FT   CARBOHYD     14     14       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    433    433       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    477    477       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    485    485       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    531    531       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    601    601       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1409   1409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1497   1497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1550   1550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1558   1558       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1678   1678       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1776   1776       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2055   2055       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2436 AA;  269971 MW;  9E6688D8615DE06D CRC64;

  Query Match             17.5%;  Score 267;  DB 1;  Length 2436;
  Best Local Similarity   29.2%;  Pred. No. 3.5e-14;
  Matches   80;  Conservative   35;  Mismatches   73;  Indels   86;  Gaps   12;

Qy      44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
           | : |:|                :| | |:|                || | ||:|| :
Db    1572 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 1631

Qy      72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
           |: : :  :          | ||:         ||   ||  ||| :  | ||| |:|
Db    1632 PRLVREPVR-----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDIT 1675

Qy     132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
           | |:|:||: |  :              |||  : |  |     |:    |   ::|
Db    1676 GHNVSEYLLFTSDRF-----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 1721

Qy     192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                                 :  |   :|::||||:|:: ::|| :||||||||| |
Db    1722 ---------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 1759

Qy     252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
           : ||: ||||  |||:| |   || :  |  | |
Db    1760 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 1793


RESULT 5
ABC2_MOUSE
ID   ABC2_MOUSE     STANDARD;      PRT;  2434 AA.
AC   P41234;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette
DE   transporter 2) (ATP-binding cassette 2).
GN   ABCA2 OR ABC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND REVISIONS.
RC   STRAIN=DBA/2;
RA   Chimini G.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 964-2434 FROM N.A.
RC   STRAIN=DBA/2; TISSUE=Macrophage;
RX   MEDLINE=94375008; PubMed=8088782;
RA   Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;
RT   "Cloning of two novel ABC transporters mapping on human chromosome
RT   9.";
RL   Genomics 21:150-159(1994).
CC   -!- FUNCTION: PROBABLE TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN
CC       FOUND YET. MAY HAVE A ROLE IN MACROPHAGE LIPID METABOLISM AND
CC       NEURAL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST
CC       LEVELS ARE FOUND IN BRAIN AND PREGNANT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X75927; CAA53531.2; -.
DR   MGD; MGI:99606; Abca2.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein.
FT   TRANSMEM     21     40       POTENTIAL.
FT   TRANSMEM    705    727       POTENTIAL.
FT   TRANSMEM    748    770       POTENTIAL.
FT   TRANSMEM    780    802       POTENTIAL.
FT   TRANSMEM    809    831       POTENTIAL.
FT   TRANSMEM   1793   1815       POTENTIAL.
FT   TRANSMEM   1846   1865       POTENTIAL.
FT   TRANSMEM   1875   1897       POTENTIAL.
FT   TRANSMEM   1904   1926       POTENTIAL.
FT   NP_BIND    1024   1031       ATP (POTENTIAL).
FT   NP_BIND    2088   2095       ATP (POTENTIAL).
FT   CARBOHYD     14     14       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    420    420       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    484    484       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    494    494       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1408   1408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1496   1496       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1549   1549       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1557   1557       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1613   1613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1678   1678       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1776   1776       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2055   2055       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2434 AA;  270582 MW;  3CEDD48ED5692005 CRC64;

  Query Match             17.3%;  Score 264;  DB 1;  Length 2434;
  Best Local Similarity   25.2%;  Pred. No. 6.3e-14;
  Matches   92;  Conservative   40;  Mismatches   97;  Indels  136;  Gaps   15;

Qy       2 GKYPSLELQPWMYNEQYT--------FVSNDAPE-------DTGTLELLNALTKDPGFGT 46
Db    1482 GDLPPLVLSPSQYH-NYTQPRGNFIPYANEERQEYRLRLSPDASPQQLVSTFRLPSGVGA 1540

Qy      47 RCM-----------------------------------EGNPI---------------- 54
Db    1541 TCVLKSPANGSLGPMLNLSSGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPAPSDS 1600

Qy      55 ---PD----------TPCQAGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKI 100
Db    1601 PVXPDEDSLQAWNMSLPPTAGPETWTSAPSLPRLVHEPVR-----------CTCSAQGT 1648

Qy     101 KKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEF 160
Db    1649 GFS---CPSSVGG-HPPQMRVVTGDILTDITGHNVSEYLLFTSDRF-----------RLH 1693

Qy     161 RYGGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
Db    1694 RYGAITFG--NVQKSIPAS----------------------FGARVPPMVRKIAVRRVA 1728

Qy     221 KVWFNNKGWHAISSFLNVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQLS-EVAL 278
```

```
Db    1729 QVLYNNKGYHSMPTYLNSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLL 1788

Qy     279 MTTSV 283
Db    1789 QGTDV 1793


RESULT 6
GLN2_DAUCA
ID   GLN2_DAUCA     STANDARD;      PRT;   432 AA.
AC   O22506;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamine synthetase, chloroplast precursor (EC 6.3.1.2) (Glutamate--
DE   ammonia ligase) (GS2).
GN   GLN2.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. US-Harumakigosun; TISSUE=Leaf;
RA   Higashi K., Kamada H.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF019561; AAB71693.1; -.
DR   InterPro; IPR001691; GLN_synth.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family; Chloroplast; Transit peptide.
FT   TRANSIT       1      ?       CHLOROPLAST (POTENTIAL).
FT   CHAIN         ?    432       GLUTAMINE SYNTHETASE.
SQ   SEQUENCE   432 AA;  47763 MW;  20EC0A4CF8E35345 CRC64;

  Query Match              7.1%;  Score 108.5;  DB 1;  Length 432;
  Best Local Similarity   24.5%;  Pred. No. 0.092;
  Matches   71;  Conservative   26;  Mismatches  110;  Indels   83;  Gaps   15;

Qy       3 KYPSLELQPWMYNEQYTFVSNDAPEDTGTLELL-NALTKDPGFG---------TRCMEGN 52
Db     105 EHPS-ELPKWNYDGSST---GQAPGDDSEVILYPQAIFKDPFRGGNNILVICDTYTPQGE 160

Qy      53 PIPDTPCQ-----------AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIK 101
Db     161 PIPTNKRHKAAQIFSDAKVLGEVPWFGIEQEYTLMQ--QDVNW---------------- 201

Qy     102 KMLPVCPPG--AGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNE 159
Db     202 ------PLGWNVGGYPGPQGPYYCAAGADKSFGRDISDAHYKACL------------- 240

Qy     160 FRYGGFSLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDT 216
```

```
Db    241 --YAGINISGTNGEVMPGQWEFQVGPSVGIEAGDHIWCAR-YLLERITEQAGVVLT-LDP 296

Qy    217 RNNVKVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
          :     | |  | |   |             ||  |||  :|:  |: | ||
Db    297 KPIDGDW-NGAGCHTNYSTKSMREEGGFEVIKKAILNLSLRHKEHISAYG 345


RESULT 7
GLN2_HORVU
ID   GLN2_HORVU     STANDARD;      PRT;   434 AA.
AC   P13564;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamine synthetase leaf isozyme, chloroplast precursor (EC 6.3.1.2)
DE   (Glutamate--ammonia ligase) (Chloroplast GS2).
OS   Hordeum vulgare (Barley).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Hordeum.
OX   NCBI_TaxID=4513;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91355850; PubMed=1983297;
RA   Stroman P., Baima S., Casadoro G.;
RT   "A cDNA sequence coding for glutamine synthetase in Hordeum vulgare
RT   L.";
RL   Plant Mol. Biol. 15:161-163(1990).
RN   [2]
RP   SEQUENCE OF 9-434 FROM N.A.
RC   STRAIN=cv. Maris Mink; TISSUE=Leaf;
RX   MEDLINE=91346618; PubMed=1983286;
RA   Freeman J., Marquez A.J., Wallsgrove R.M., Saarelainen R.,
RA   Forde B.G.;
RT   "Molecular analysis of barley mutants deficient in chloroplast
RT   glutamine synthetase.";
RL   Plant Mol. Biol. 14:297-311(1990).
RN   [3]
RP   SEQUENCE OF 48-434 FROM N.A.
RX   MEDLINE=89322552; PubMed=2473765;
RA   Baima S., Haegi A., Stroman P., Casadoro G.;
RT   "Characterization of a cDNA clone for barley leaf glutamine
RT   synthetase.";
RL   Carlsberg Res. Commun. 54:1-9(1989).
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION.
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- MISCELLANEOUS: IN BARLEY, THERE ARE DISTINCT ISOZYMES IN THE
CC       CHLOROPLAST, AND CYTOPLASM.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53580; CAA37643.1; -.
DR   EMBL; X16000; CAA34131.1; -.
DR   PIR; S11865; AJBHQ.
DR   InterPro; IPR001691; GLN_synth.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family; Chloroplast; Transit peptide.
FT   TRANSIT       1     54       CHLOROPLAST.
FT   CHAIN        55    434       GLUTAMINE SYNTHETASE LEAF ISOZYME.
FT   CONFLICT      9     27       GAGGCAGDAVPGGGEGQDG ->
FT                                AQAVVQAMQCQVGVRGRTA (IN REF. 2).
FT   CONFLICT     41     41       S -> R (IN REF. 2).
SQ   SEQUENCE   434 AA;  47094 MW;  FC47F5685EFC0D1E CRC64;

  Query Match             6.5%;  Score 98.5;  DB 1;  Length 434;
  Best Local Similarity   23.8%;  Pred. No. 0.65;
  Matches   68;  Conservative   29;  Mismatches  110;  Indels   79;  Gaps   14;

Qy      5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
          || ||  | |:   |     || ||:  :    |: |||  |         |   :| ||
Db    109 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVICDTYTPQGEPI 164

Qy     55 PDT----PCQAGEEEWTTAPVP-------QTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
          |      |   :   |: ||         |:|   :: || :  |
Db    165 PTNKRHMAAQIFSDPKVTSQVPWFGIEQEYTLMQ--RDVNWPLGWP-------------- 208

Qy    104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISD--YLVKTYVQIIAKSLKNKIWVNEFR 161
                || | ||     |       ||:||| |    | |       ::   ::
Db    209 -------VGGYPGPQGPYYCAVGSDKSFGRDISDAHYKACLYAGIEISGTNGEVMPGQWE 261

Qy    162 YG-GFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
          |  | |:|:     :  |: :                    :|     |  :| || :
Db    262 YQVGPSVGIDAGDHIWASRYI-----------------LERITEQAGVVLT-LDPKPIQ 302

Qy    221 KVWFNNKGWHAISSFL--------NVINNAILRANLQKGENPSHYG 258
            | |  | |   | |         :||  |||  :|:   : : ||
Db    303 GDW-NGAGCHTNYSTLSMREDGGFDVIKKAILNLSLRHDLHIAAYG 347


RESULT 8
DYSF_MOUSE
ID   DYSF_MOUSE     STANDARD;      PRT;  2083 AA.
AC   Q9ESD7; Q9QXC0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dysferlin (Dystrophy associated fer-1 like protein) (Fer-1 like
DE   protein 1).
GN   DYSF OR FER1L1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c, C57BL/10, and SJL/J; TISSUE=Skeletal muscle;
RX   MEDLINE=21129430; PubMed=11234777;
RA   Vafiadaki E., Reis A., Keers S., Harrison R., Anderson L.V.B.,
RA   Raffelsberger T., Ivanova S., Hoeger H., Bittner R.E., Bushby K.M.D.,
RA   Bashir R.;
RT   "Cloning of the mouse dysferlin gene and genomic characterization of
RT   the SJL-Dysf mutation.";
RL   NeuroReport 12:625-629(2001).
RN   [2]
RP   SEQUENCE OF 468-2083 FROM N.A.
RC   STRAIN=BALB/c, C57BL/10, C57BL/6, C3H, and B6C3FE;
RX   MEDLINE=99438383; PubMed=10508505;
RA   Bittner R.E., Anderson L.V.B., Burkhardt E., Bashir R., Vafiadaki E.,
RA   Ivanova S., Raffelsberger T., Maerk I., Hoeger H., Jung M.,
RA   Karbasiyan M., Storch M., Lassmann H., Moss J.A., Davison K.,
RA   Harrison R., Bushby K.M.D., Reis A.;
RT   "Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for
RT   limb girdle muscular dystrophy 2B.";
RL   Nat. Genet. 23:141-142(1999).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Localizes to the
CC       sarcolemma (By similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in skeletal and cardiac muscle. Also
CC       found in brain, liver, and kidney.
CC   -!- DISEASE: Defects in Dysf are the cause of a slowly progressive
CC       muscular dystrophy observed in SJL mice. It affects primarily the
CC       proximal muscles and it is inherited as autosomal recessive trait.
CC   -!- SIMILARITY: BELONGS TO THE FERLIN FAMILY.
```

```
CC   -!- SIMILARITY: CONTAINS 5 C2 DOMAINS.
CC   -!- CAUTION: The submitted sequence from Ref.1 differs from that shown
CC       in a number of positions. We use the sequence obtained by direct
CC       sequencing which is cited in figure 1 of that reference.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF188290; AAG17046.2; ALT_SEQ.
DR   EMBL; AJ242954; CAB63111.1; -.
DR   HSSP; P04410; 1A25.
DR   MGD; MGI:1349385; Dysf.
DR   InterPro; IPR000008; C2.
DR   Pfam; PF00168; C2; 7.
DR   SMART; SM00239; C2; 7.
DR   PROSITE; PS00499; C2_DOMAIN_1; FALSE_NEG.
DR   PROSITE; PS50004; C2_DOMAIN_2; 5.
KW   Transmembrane; Repeat.
FT   DOMAIN       1   2049       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  2050   2070       POTENTIAL.
FT   DOMAIN    2071   2083       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       1     86       C2 DOMAIN 1.
FT   DOMAIN     210    305       C2 DOMAIN 2.
FT   DOMAIN     369    487       C2 DOMAIN 3.
FT   DOMAIN    1142   1247       C2 DOMAIN 4.
FT   DOMAIN    1568   1666       C2 DOMAIN 5.
FT   DOMAIN    1041   1100       ARG-RICH.
FT   CONFLICT   468    468       D -> G (IN REF. 2).
FT   CONFLICT   502    509       AVKPSKAS -> VLKSPQAT (IN REF. 2).
FT   CONFLICT  1687   1687       C -> Y (IN REF. 2).
FT   CONFLICT  1694   1694       K -> Q (IN REF. 2).
FT   CONFLICT  1698   1698       K -> Q (IN REF. 2).
FT   CONFLICT  1799   1799       Q -> R (IN REF. 2).
FT   CONFLICT  1847   1853       MVGFEEH -> ISGSEEN (IN REF. 2).
FT   CONFLICT  1865   1865       G -> D (IN REF. 2).
FT   CONFLICT  1886   1888       VCA -> LCI (IN REF. 2).
FT   CONFLICT  1893   1894       DA -> EH (IN REF. 2).
FT   CONFLICT  1897   1911       RLDKTESKIPARVVF -> SIDQTEFRVPPRLII (IN
FT                               REF. 2).
FT   CONFLICT  1921   1921       F -> L (IN REF. 2).
FT   CONFLICT  1924   1924       F -> Y (IN REF. 2).
FT   CONFLICT  1927   1929       SLQ -> FLE (IN REF. 2).
FT   CONFLICT  1933   1937       NRMPK -> HRTII (IN REF. 2).
FT   CONFLICT  1942   1942       A -> S (IN REF. 2).
FT   CONFLICT  1949   1960       QLDDTFHPEWFV -> MIPDLKAMDPLKAKTA (IN REF.
FT                               2).
FT   CONFLICT  1966   1968       KTV -> RSM (IN REF. 2).
FT   CONFLICT  1975   1984       VTEEGEKKML -> YADKDGTRVM (IN REF. 2).
FT   CONFLICT  1988   1988       L -> V (IN REF. 2).
FT   CONFLICT  1994   2001       IVAESEHE -> VLNEREAD (IN REF. 2).
FT   CONFLICT  2007   2007       Q -> K (IN REF. 2).
FT   CONFLICT  2010   2010       D -> S (IN REF. 2).
FT   CONFLICT  2019   2025       EDPRRPD -> DPPNRPE (IN REF. 2).
FT   CONFLICT  2033   2035       SPY -> NPC (IN REF. 2).
FT   CONFLICT  2039   2039       K -> R (IN REF. 2).
FT   CONFLICT  2042   2042       L -> V (IN REF. 2).
FT   CONFLICT  2047   2057       RCAIILFIILF -> KWVIIGLLLLL (IN REF. 2).
FT   CONFLICT  2063   2070       LGVFVYAF -> VAVLLYSL (IN REF. 2).
FT   CONFLICT  2074   2075       AA -> LS (IN REF. 2).
FT   CONFLICT  2078   2083       LVKPFR -> IVRPNA (IN REF. 2).
SQ   SEQUENCE  2083 AA;  237127 MW;  833C5A8FC5673ABA CRC64;

  Query Match             6.5%;  Score 98.5;  DB 1;  Length 2083;
  Best Local Similarity   18.5%;  Pred. No. 5.5;
  Matches   67;  Conservative   43;  Mismatches   95;  Indels  157;  Gaps   15;

Qy       1 FGKYPSL--------ELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGN 52
             |||   |        :|:  :|:  |  :| |       ::| | |     || ||

Db    1630 FGKMFELTCTLPLEKDLKITLYD--YDLLSKDEKIGETVIDLENRLLSK--FGARC---- 1681

Qy      53 PIPDTPCQAGEEEW-----------------TTAPVPQTIMDLFQNGNWTMQN------ 88
            :| | | :|  :|                     ||| :|    ||: ::|::
Db    1682 GLPQTCCVSGPNKWRDKLRPSQLLHLFCQQHRIKAPVYRTDRVTFQDKDYTIEEIEAGRL 1741

Qy      89 PSPAC---------------------------------QCSSDKIKKMLPVCPPGAGGLP 115
           |:|                                       |::  : : |   |
Db    1742 PNPHLGPVEERLALHVLQQQGLVPEHVESRPLYSPLQPDIEQGKLQMWIDIFPKVLGQPG 1801

Qy     116 PP----QRKQ----------NTADILQD---LTGRNISDYLVKTYV-------------- 144
           ||     ||           || |:: |   |||  :|| || ::
Db    1802 PPFNITPRKARRFFLRCIIWNTKDVILDDLSLTGEKMSDIYVKGWMVGFEEHKQKTDVHY 1861

Qy     145 ----------------------QIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEV 181
                                  |: | : |:  |            :  |::  |:: |
Db    1862 RSLGGEGNFNWRFVFPFDYLPAEQVCAVAKKDAFW----------RLDKTESKIPARVV 1910

Qy     182 -----NDA-------------IKQMKKHLKLAKDSSADRFLNSL-GRFMTGLDTRNNVKV 222
                 ||              : :| |  | |:  | |:  ::    :  |  :  ||
Db    1911 FQIWDNDKFSFDDFLGSLQLDLNRMPKPAKTAEKCSLDQLDDTFHPEWFVSLFEQKTVKG 1970

Qy     223 WF 224
           |:
Db    1971 WW 1972


RESULT 9
GLN4_PHAVU
ID   GLN4_PHAVU     STANDARD;      PRT;   429 AA.
AC   P15102;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamine synthetase leaf isozyme, chloroplast precursor (EC 6.3.1.2)
DE   (Isozyme delta) (Glutamate--ammonia ligase).
OS   Phaseolus vulgaris (Kidney bean) (French bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX   NCBI_TaxID=3885;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Tendergreen; TISSUE=Leaf;
RA   Lightfoot D.A., Green N.K., Cullimore J.V.;
RT   "The chloroplast-located glutamine synthetase of Phaseolus vulgaris
RT   L.: nucleotide sequence, expression in different organs and uptake
RT   into isolated chloroplasts.";
RL   Plant Mol. Biol. 11:191-202(1988).
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION.
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- MISCELLANEOUS: THERE ARE AT LEAST FOUR ISOZYMES OF THIS ENZYME IN
CC       P.VULGARIS.
CC   -!- MISCELLANEOUS: IRREVERSIBLE INHIBITED BY THE HERBICIDE
CC       L-PHOSPHINOTHRICIN (PPT).
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12738; CAA31234.1; -.
DR   PIR; S04031; AJFBQD.
DR   InterPro; IPR001691; GLN_synth.
```

```
DR   InterPro; IPR001637; GlnA_adenyltn.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Nitrogen fixation; Ligase; Multigene family; Chloroplast;
KW   Transit peptide.
FT   TRANSIT       1     57       CHLOROPLAST.
FT   CHAIN        58    429       GLUTAMINE SYNTHETASE LEAF ISOZYME.
SQ   SEQUENCE   429 AA;  47246 MW;  0CA55624B1118AF8 CRC64;

  Query Match             6.3%;  Score 95.5;  DB 1;  Length 429;
  Best Local Similarity   24.1%;  Pred. No. 1.2;
  Matches   69;  Conservative   26;  Mismatches  116;  Indels   75;  Gaps   14;

Qy       3 KYPSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFGTRCM---------EGN 52
           ::|| ||  | |:   |     || ||:  :    |: |||  |   :          |
Db     102 EHPS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVICDAYTPAGE 157

Qy      53 PIPDTPCQAGEEEWTT----APVP-----QTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
           |||        | ::     | ||     |    |  | || :  |
Db     158 PIPTNKRHRAAEVFSNPRVIAEVPWFGIEQEYTLLQTNVNWPLGWP-------------- 203

Qy     104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISD--YLVKTYVQIIAKSLKNKIWVNEFR 161
                 || | ||     :       ||:|||  |    :  |       ::   ::
Db     204 -------VGGYPGPQGPYYCSAGADKSFGRDISDAHYKACLFAGINISGTNGEVMPGQWE 256

Qy     162 YG-GFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNV 220
           .|  | |:|:          |  | |    :  |        :|     |  :  || :
Db     257 YQVGPSVGI----------EAGDHIWASRYIL--------ERITEQAG-VVLSLDPKPIE 297

Qy     221 KVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
           { |  | |   |           ||  |||  :|:  |: | ||
Db     298 GDW-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHKEHISAYG 342


RESULT 10
Y314_METJA
ID   Y314_METJA     STANDARD;      PRT;   263 AA.
AC   Q57762;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0314.
GN   MJ0314.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: SOME, TO M.JANNASCHII MJ0398.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67486; AAB98310.1; -.
DR   TIGR; MJ0314; -.
DR   InterPro; IPR002197; HTH_Fis.
DR   InterPro; IPR000792; HTH_LuxR.
DR   InterPro; IPR004042; Intein_endonuc.
DR   PROSITE; PS50819; INTEIN_ENDONUCLEASE; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   263 AA;  30804 MW;  A7520A3BBE0CC5CD CRC64;

  Query Match             6.2%;  Score 94.5;  DB 1;  Length 263;
  Best Local Similarity   24.6%;  Pred. No. 0.73;
  Matches   49;  Conservative   32;  Mismatches   61;  Indels   57;  Gaps   11;

Qy      72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
           || |: |:||| :|    :   :|| : |:::|              : |  ||
Db      43 PQEIIKLYQNG-YTTTEIAKIMKCSHETIRRIL--------------RNNNIDI----- 81

Qy     132 GRNISDYLVKTYVQIIAKSLKN--KIWVNEFRYGGFSLGVSNTQALPPSQEVN------- 182
            |  |: |:          :||  || :|      : ||| |       || |
Db      82 -RKSSESLI----------IKNTKKINLNPSESLAYILGVLNGDGSVNKQESNYVIELKV 130

Qy     183 ---DAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKG---WHA---IS 233
             | |:: |::|    ::   :::|   |       :  |:|   :||   |:    :
Db     131 TDKDFIEEFKRNL----ENIGFKYINEYVRKFENKKDQYVVRV--RSKGFYYWYKSLNVD 184

Qy     234 SFLNVI-NNAILRANLQKG 251
           ::||| ||  |  :  ||
Db     185 YYMNVIGNNEKLMISWLKG 203


RESULT 11
GLN2_ARATH
ID   GLN2_ARATH     STANDARD;      PRT;   430 AA.
AC   Q43127;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamine synthetase, chloroplast precursor (EC 6.3.1.2) (Glutamate--
DE   ammonia ligase) (GS2).
GN   GLN2 OR GSLI OR AT5G35630 OR MJE4.9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92079889; PubMed=1684022;
RA   Peterman T.K., Goodman H.M.;
RT   "The glutamine synthetase gene family of Arabidopsis thaliana: light-
RT   regulation and differential expression in leaves, roots and seeds.";
RL   Mol. Gen. Genet. 230:145-154(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Arimura G., Fujii M., Takahashi M., Goshima N., Morikawa H.;
RT   "Nucleotide sequences of genes for cytosolic and chloroplastic
RT   glutamine synthetase from Arabidopsis thaliana.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=98403884; PubMed=9734815;
RA   Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT   Sequence features of the regions of 1,367,185 bp covered by 19
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:203-216(1998).
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
```

```
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S69727; AAB20558.1; -.
DR   EMBL; AB015045; BAA88761.1; -.
DR   EMBL; AB013393; BAB09304.1; -.
DR   SWISS-2DPAGE; Q43127; ARATH.
DR   InterPro; IPR001691; GLN_synth.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family; Chloroplast; Transit peptide.
FT   TRANSIT       1     51       CHLOROPLAST (POTENTIAL).
FT   CHAIN        52    430       GLUTAMINE SYNTHETASE.
SQ   SEQUENCE   430 AA;  47410 MW;  664029BC06572295 CRC64;

  Query Match             6.2%;  Score 94.5;  DB 1;  Length 430;
  Best Local Similarity   23.1%;  Pred. No. 1.4;
  Matches   67;  Conservative   25;  Mismatches  111;  Indels   87;  Gaps   14;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
           || ||  | |:   |     || ||:  :    |: :||  |          |    | ||
Db     105 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFRDPFRGGNNILVICDTWTPAGEPI 160

Qy      55 P-----------DTPCQAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKM 103
           |                :||  |       |::   ||  | :  |
Db     161 PTNKRAKAAEIFSNKKVSGEVPWFGIEQEYTLLQ--QNVKWPLGWP-------------- 204

Qy     104 LPVCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYG 163
                 |  | ||          : ||:|||   |  :                  |
Db     205 -------VGAFPGPQGPYYCGVGADKIWGRDISDAHYKACL----------------YA 240

Qy     164 GFSLGVSNTQALPPSQEVN-------DAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDT 216
           | ::  :| : :|   |          ||      |:  |:    :|     |  :| ||
Db     241 GINISGTNGEVMPGQWEFQVGPSVGIDA----GDHVWCAR-YLLERITEQAGVVLT-LDP 294

Qy     217 RNNVKVWFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
           :    | |  | |   |           ||  |||  :|:  |: | ||
Db     295 KPIEGDW-NGAGCHTNYSTKSMREEGGFEVIKKAILNLSLRHKEHISAYG 343


RESULT 12
DYSF_HUMAN
ID   DYSF_HUMAN     STANDARD;      PRT;  2080 AA.
AC   O75923; O75696; Q9UEN7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dysferlin (Dystrophy associated fer-1-like protein) (Fer-1 like
DE   protein 1).
GN   DYSF OR FER1L1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., VARIANTS MM V-1298; R-1857 AND C-2042, AND
RP   VARIANTS LGMD2B V-1298 AND C-2042.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=98400252; PubMed=9731526;
RA   Liu J., Aoki M., Illa I., Wu C., Fardeau M., Angelini C., Serrano C.,
RA   Urtizberea J.A., Hentati F., Hamida M.B., Bohlega S., Culper E.J.,
RA   Amato A.A., Bossie K., Oeltjen J., Bejaoui K., McKenna-Yasek D.,
RA   Hosler B.A., Schurr E., Arahata K., de Jong P.J., Brown R.H. Jr.;
RT   "Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi
RT   myopathy and limb girdle muscular dystrophy.";
RL   Nat. Genet. 20:31-36(1998).
RN   [2]
RP   SEQUENCE OF 303-2080 FROM N.A.
RC   TISSUE=Skeletal muscle, and Placenta;
RX   MEDLINE=98400253; PubMed=9731527;
RA   Bashir R., Britton S., Strachan T., Keers S., Vafiadaki E., Lako M.,
RA   Richard I., Marchand S., Bourg N., Argov Z., Sadeh M., Mahjneh I.,
RA   Marconi G., Passos-Bueno M.R., de Sa Moreira E., Zatz M.,
RA   Beckmann J.S., Bushby K.M.D.;
RT   "A gene related to Caenorhabditis elegans spermatogenesis factor fer-1
RT   is mutated in limb-girdle muscular dystrophy type 2B.";
RL   Nat. Genet. 20:37-42(1998).
RN   [3]
RP   SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=99214026; PubMed=10196375;
RA   Anderson L.V.B., Davison K., Moss J.A., Young C., Cullen M.J.,
RA   Walsh J., Johnson M.A., Bashir R., Britton S., Keers S., Argov Z.,
RA   Mahjneh I., Fougerousse F., Beckmann J.S., Bushby K.M.D.;
RT   "Dysferlin is a plasma membrane protein and is expressed early in
RT   human development.";
RL   Hum. Mol. Genet. 8:855-861(1999).
RN   [4]
RP   ERRATUM.
RA   Anderson L.V.B., Davison K., Moss J.A., Young C., Cullen M.J.,
RA   Walsh J., Johnson M.A., Bashir R., Britton S., Keers S., Argov Z.,
RA   Mahjneh I., Fougerousse F., Beckmann J.S., Bushby K.M.D.;
RL   Hum. Mol. Genet. 8:1141-1141(1999).
RN   [5]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=99424596; PubMed=10496277;
RA   Matsuda C., Aoki M., Hayashi Y.K., Ho M.F., Arahata K.,
RA   Brown R.H. Jr.;
RT   "Dysferlin is a surface membrane-associated protein that is absent in
RT   Miyoshi myopathy.";
RL   Neurology 53:1119-1122(1999).
RN   [6]
RP   SUBCELLULAR LOCATION, AND VARIANT MM AND LGMD2B ARG-791.
RX   MEDLINE=99214028; PubMed=10196377;
RA   Weiler T., Bashir R., Anderson L.V.B., Davison K., Moss J.A.,
RA   Britton S., Nylen E., Keers S., Vafiadaki E., Greenberg C.R.,
RA   Bushby K.M.D., Wrogemann K.;
RT   "Identical mutation in patients with limb girdle muscular dystrophy
RT   type 2B or Miyoshi myopathy suggests a role for modifier gene(s).";
RL   Hum. Mol. Genet. 8:871-877(1999).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC       Localizes to the sarcolemma.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKELETAL MUSCLE. ALSO
CC       FOUND IN HEART, PLACENTA AND AT LOWER LEVELS IN LIVER, LUNG,
CC       KIDNEY AND PANCREAS.
CC   -!- DEVELOPMENTAL STAGE: Expression in limb tissue from 5-6 weeks
CC       embryos; persists throughout development.
CC   -!- DISEASE: DEFECTS IN DYSF ARE THE CAUSE OF AUTOSOMAL RECESSIVE LIMB
CC       GIRDLE MUSCULAR DYSTROPHY TYPE 2B (LGMD2B). TYPE 2 LIMB GIRDLE
CC       MUSCULAR DYSTROPHIES REPRESENT A GENETICALLY HETEROGENEOUS GROUP
CC       OF DISEASES WITH VARYING DEGREES OF SEVERITY DEPENDING ON AGE AT
CC       ONSET AND RATE OF PROGRESSION. LGMD2B IS CHARACTERIZED BY WEAKNESS
CC       AND ATROPHY STARTING IN THE PROXIMAL PELVIFEMORAL MUSCLES, WITH
CC       ONSET IN THE LATE TEENS OR LATER, MASSIVE ELEVATION OF SERUM
CC       CREATINE KINASE LEVELS AND SLOW PROGRESSION. SCAPULAR MUSCLE
CC       INVOLVEMENT IS MINOR AND NOT PRESENT AT ONSET. UPPER LIMB GIRDLE
CC       INVOLVEMENT FOLLOWS SOME YEARS AFTER THE ONSET IN LOWER LIMBS.
CC   -!- DISEASE: DEFECTS IN DYSF ARE THE CAUSE OF MIYOSHI MYOPATHY (MM).
CC       THIS TYPE OF AUTOSOMAL RECESSIVE MUSCULAR DYSTROPHY INVOLVES THE
CC       DISTAL LOWER LIMB MUSCULATURE. IT IS CHARACTERIZED BY WEAKNESS
CC       THAT INITIALLY AFFECTS THE GASTROCNEMIUS MUSCLE DURING EARLY
CC       ADULTHOOD. OTHERWISE THE PHENOTYPE OVERLAPS WITH LGMD2B,
CC       ESPECIALLY IN AGE AT ONSET AND CREATINE KINASE ELEVATION.
CC   -!- SIMILARITY: BELONGS TO THE FERLIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 C2 DOMAINS.
CC   -!- DATABASE: NAME=Leiden Muscular Dystrophy pages; NOTE=Dysferlin;
```

```
CC       WWW="http://www.dmd.nl/dysf_home.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF075575; AAC63519.1; -.
DR   EMBL; AJ007670; CAA07603.1; ALT_SEQ.
DR   EMBL; AJ007973; CAA07800.1; -.
DR   HSSP; P21707; 1RSY.
DR   Genew; HGNC:3097; DYSF.
DR   MIM; 603009; -.
DR   MIM; 253601; -.
DR   MIM; 254130; -.
DR   MIM; 606768; -.
DR   InterPro; IPR000008; C2.
DR   Pfam; PF00168; C2; 7.
DR   SMART; SM00239; C2; 7.
DR   PROSITE; PS00499; C2_DOMAIN_1; FALSE_NEG.
DR   PROSITE; PS50004; C2_DOMAIN_2; 5.
KW   Transmembrane; Repeat; Disease mutation.
FT   DOMAIN       1   2046       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  2047   2067       POTENTIAL.
FT   DOMAIN    2068   2080       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       1     85       C2 DOMAIN 1.
FT   DOMAIN     207    302       C2 DOMAIN 2.
FT   DOMAIN     366    479       C2 DOMAIN 3.
FT   DOMAIN    1139   1244       C2 DOMAIN 4.
FT   DOMAIN    1565   1663       C2 DOMAIN 5.
FT   DOMAIN    1038   1097       ARG-RICH.
FT   VARIANT    791    791       P -> R (IN MM AND LGMD2B).
FT                               /FTId=VAR_012308.
FT   VARIANT   1298   1298       I -> V (IN MM AND LGMD2B).
FT                               /FTId=VAR_012309.
FT   VARIANT   1857   1857       H -> R (IN MM).
FT                               /FTId=VAR_012310.
FT   VARIANT   2042   2042       R -> C (IN MM AND LGMD2B).
FT                               /FTId=VAR_012311.
SQ   SEQUENCE   2080 AA;  237293 MW;  376E25A5AB9BE398 CRC64;

  Query Match             6.2%;  Score 94.5;  DB 1;  Length 2080;
  Best Local Similarity   18.0%;  Pred. No. 12;
  Matches   65;  Conservative   45;  Mismatches   95;  Indels  157;  Gaps   15;

Qy       1 FGKYPSL--------ELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGN 52
           |||   |        :|:  :|:  |  :| |       ::| | |     || ||
Db    1627 FGKMFELTCTLPLEKDLKITLYD--YDLLSKDEKIGETVVDLENRLLSK--FGARC---- 1678

Qy      53 PIPDTPCQAGEEEW-----------------TTAPVPQTIMDLFQNGNWTMQN------ 88
            :| | | :|  :|                  ||| :|   :||:  ::::
Db    1679 GLPQTYCVSGPNQWRDQLRPSQLLHLFCQQHRVKAPVYRTDRVMFQDKEYSIEEIEAGRI 1738

Qy      89 PSPAC---------------------------------QCSSDKIKKMLPVCPPGAGGLP 115
           |:|                                    |::  : : |   |
Db    1739 PNPHLGPVEERLALHVLQQQGLVPEHVESRPLYSPLQPDIEQGKLQMWVDLFPKALGRPG 1798

Qy     116 P-----PQRKQ---------NTADILQD---LTGRNISDYLVKTYV-------------- 144
           |     |:| :          || |:: |   |||  :|| || ::
Db    1799 PPFNITPRRARRFFLRCIIWNTRDVILDDLSLTGEKMSDIYVKGWMIGFEEHKQKTDVHY 1858

Qy     145 ----------------------QIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEV 181
                               |:   : |:  |           :  |::  |:: |
Db    1859 RSLGGEGNFNWRFIFPFDYLPAEQVCTIAKKDAFW-----------RLDKTESKIPARVV 1907

Qy     182 -----NDA-------------IKQMKKHLKLAKDSSADRFLNSL-GRFMTGLDTRNNVKV 222
                ||              : :| |  | ||  | |:  ::    :   |  :  ||
Db    1908 FQIWDNDKFSFDDFLGSLQLDLNRMPKPAKTAKKCSLDQLDDAFHPEWFVSLFEQKTVKG 1967

Qy     223 WF 224
           |:
Db    1968 WW 1969


RESULT 13
LRP2_RAT
ID   LRP2_RAT       STANDARD;      PRT;  4660 AA.
AC   P98158;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Low-density lipoprotein receptor-related protein 2 precursor (Megalin)
DE   (Glycoprotein 330).
GN   LRP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX   MEDLINE=95024033; PubMed=7937880;
RA   Saito A., Pietromonaco S., Loo A.K.C., Farquhar M.G.;
RT   "Complete cloning and sequencing of rat gp330/'megalin,' a
RT   distinctive member of the low density lipoprotein receptor gene
RT   family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:9725-9729(1994).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=95386696; PubMed=7544804;
RA   Moestrup S.K., Cui S., Vorum H., Bregengaard C., Bjorn S.E.,
RA   Norris K., Gliemann J., Christensen E.I.;
RT   "Evidence that epithelial glycoprotein 330/megalin mediates uptake of
RT   polybasic drugs.";
RL   J. Clin. Invest. 96:1404-1413(1995).
RN   [3]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=94172242; PubMed=7510321;
RA   Zheng G., Bachinsky D.R., Stamenkovic I., Strickland D.K., Brown D.,
RA   Andres G., McCluskey R.T.;
RT   "Organ distribution in rats of two members of the low-density
RT   lipoprotein receptor gene family, gp330 and LRP/alpa 2MR, and the
RT   receptor-associated protein (RAP).";
RL   J. Histochem. Cytochem. 42:531-542(1994).
CC   -!- FUNCTION: BINDS PLASMINOGEN, EXTRACELLULAR MATRIX COMPONENTS,
CC       PLASMINOGEN ACTIVATOR-PLASMINOGEN ACTIVATOR INHIBITOR TYPE I
CC       COMPLEX, APOLIPOPROTEIN E-ENRICHED BETA-VLDL, LIPOPROTEIN LIPASE,
CC       LACTOFERRIN, CLUSTERIN AND CALCIUM.
CC   -!- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS
CC       APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B.
CC   -!- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-
CC       ASSOCIATED PROTEIN (RAP).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXPRESSED IN
CC       CLATHRIN-COATED PITS; A SOLUBLE FORM IS POSSIBLY DERIVED BY
CC       CLEAVAGE AT THE CELL SURFACE.
CC   -!- TISSUE SPECIFICITY: EPITHELIAL CELLS OF KIDNEY GLOMERULUS AND
CC       PROXIMAL TUBULE, LUNG, EPIDIDYMIS, YOLK SAC, AMONG OTHERS.
CC   -!- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.
CC   -!- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34049; AAA51369.1; -.
DR   HSSP; Q07954; 1CR8.
DR   GlycoSuiteDB; P98158; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
```

```
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000033; Ldl_receptor_rep.
DR   Pfam; PF00057; ldl_recept_a; 36.
DR   Pfam; PF00058; ldl_recept_b; 33.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00001; EGF_like; 15.
DR   SMART; SM00192; LDLa; 36.
DR   SMART; SM00135; LY; 35.
DR   PROSITE; PS00010; ASX_HYDROXYL; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 8.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01209; LDLRA_1; 31.
DR   PROSITE; PS50068; LDLRA_2; 36.
KW   Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane;
KW   Receptor; EGF-like domain; SH3-binding; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   4660       LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
FT                                PROTEIN 2.
FT   DOMAIN      26   4425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  4426   4446       POTENTIAL.
FT   DOMAIN    4447   4660       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      26     64       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      65    105       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN     106    144       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN     145    181       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN     181    219       LDL-RECEPTOR CLASS A 5.
FT   DOMAIN     220    258       LDL-RECEPTOR CLASS A 6.
FT   DOMAIN     263    307       LDL-RECEPTOR CLASS A 7.
FT   DOMAIN     308    346       EGF-LIKE 1.
FT   DOMAIN     347    385       EGF-LIKE 2.
FT   DOMAIN     435    476       LDL-RECEPTOR CLASS B 1.
FT   DOMAIN     478    519       LDL-RECEPTOR CLASS B 2.
FT   DOMAIN     521    566       LDL-RECEPTOR CLASS B 3.
FT   DOMAIN     568    611       LDL-RECEPTOR CLASS B 4.
FT   DOMAIN     612    652       LDL-RECEPTOR CLASS B 5.
FT   DOMAIN     658    704       EGF-LIKE 3.
FT   DOMAIN     752    793       LDL-RECEPTOR CLASS B 6.
FT   DOMAIN     795    835       LDL-RECEPTOR CLASS B 7.
FT   DOMAIN     837    879       LDL-RECEPTOR CLASS B 8.
FT   DOMAIN     881    923       LDL-RECEPTOR CLASS B 9.
FT   DOMAIN     969   1013       EGF-LIKE 4.
FT   DOMAIN    1023   1061       LDL-RECEPTOR CLASS A 8.
FT   DOMAIN    1064   1103       LDL-RECEPTOR CLASS A 9.
FT   DOMAIN    1108   1146       LDL-RECEPTOR CLASS A 10.
FT   DOMAIN    1148   1186       LDL-RECEPTOR CLASS A 11.
FT   DOMAIN    1186   1225       LDL-RECEPTOR CLASS A 12.
FT   DOMAIN    1229   1269       LDL-RECEPTOR CLASS A 13.
FT   DOMAIN    1270   1308       LDL-RECEPTOR CLASS A 14.
FT   DOMAIN    1311   1351       LDL-RECEPTOR CLASS A 15.
FT   DOMAIN    1350   1390       EGF-LIKE 5.
FT   DOMAIN    1391   1430       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1479   1520       LDL-RECEPTOR CLASS B 10.
FT   DOMAIN    1522   1563       LDL-RECEPTOR CLASS B 11.
FT   DOMAIN    1567   1609       LDL-RECEPTOR CLASS B 12.
FT   DOMAIN    1611   1654       LDL-RECEPTOR CLASS B 13.
FT   DOMAIN    1656   1696       LDL-RECEPTOR CLASS B 14.
FT   DOMAIN    1701   1742       EGF-LIKE 7.
FT   DOMAIN    1791   1832       LDL-RECEPTOR CLASS B 15.
FT   DOMAIN    1834   1882       LDL-RECEPTOR CLASS B 16.
FT   DOMAIN    1884   1930       LDL-RECEPTOR CLASS B 17.
FT   DOMAIN    1932   1972       LDL-RECEPTOR CLASS B 18.
FT   DOMAIN    1973   2013       LDL-RECEPTOR CLASS B 19.
FT   DOMAIN    2019   2060       EGF-LIKE 8.
FT   DOMAIN    2108   2156       LDL-RECEPTOR CLASS B 20.
FT   DOMAIN    2158   2201       LDL-RECEPTOR CLASS B 21.
FT   DOMAIN    2203   2245       LDL-RECEPTOR CLASS B 22.
FT   DOMAIN    2247   2289       LDL-RECEPTOR CLASS B 23.
FT   DOMAIN    2291   2332       LDL-RECEPTOR CLASS B 24.
FT   DOMAIN    2343   2384       EGF-LIKE 9.
FT   DOMAIN    2432   2477       LDL-RECEPTOR CLASS B 25.
FT   DOMAIN    2479   2518       LDL-RECEPTOR CLASS B 26.
FT   DOMAIN    2520   2562       LDL-RECEPTOR CLASS B 27.
FT   DOMAIN    2564   2604       LDL-RECEPTOR CLASS B 28.
FT   DOMAIN    2605   2647       LDL-RECEPTOR CLASS B 29.
FT   DOMAIN    2652   2694       EGF-LIKE 10.
FT   DOMAIN    2699   2739       LDL-RECEPTOR CLASS A 16.
FT   DOMAIN    2740   2778       LDL-RECEPTOR CLASS A 17.
FT   DOMAIN    2779   2820       LDL-RECEPTOR CLASS A 18.
FT   DOMAIN    2821   2862       LDL-RECEPTOR CLASS A 19.
FT   DOMAIN    2863   2903       LDL-RECEPTOR CLASS A 20.
FT   DOMAIN    2906   2947       LDL-RECEPTOR CLASS A 21.
FT   DOMAIN    2948   2992       LDL-RECEPTOR CLASS A 22.
FT   DOMAIN    2993   3031       LDL-RECEPTOR CLASS A 23.
FT   DOMAIN    3032   3072       LDL-RECEPTOR CLASS A 24.
FT   DOMAIN    3075   3112       LDL-RECEPTOR CLASS A 25.
FT   DOMAIN    3113   3153       EGF-LIKE 11.
FT   DOMAIN    3154   3194       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    3241   3282       LDL-RECEPTOR CLASS B 30.
FT   DOMAIN    3284   3333       LDL-RECEPTOR CLASS B 31.
FT   DOMAIN    3335   3377       LDL-RECEPTOR CLASS B 32.
FT   DOMAIN    3379   3420       LDL-RECEPTOR CLASS B 33.
FT   DOMAIN    3421   3461       LDL-RECEPTOR CLASS B 34.
FT   DOMAIN    3467   3511       EGF-LIKE 13.
FT   DOMAIN    3512   3552       LDL-RECEPTOR CLASS A 26.
FT   DOMAIN    3553   3593       LDL-RECEPTOR CLASS A 27.
FT   DOMAIN    3594   3634       LDL-RECEPTOR CLASS A 28.
FT   DOMAIN    3635   3675       LDL-RECEPTOR CLASS A 29.
FT   DOMAIN    3678   3718       LDL-RECEPTOR CLASS A 30.
FT   DOMAIN    3719   3758       LDL-RECEPTOR CLASS A 31.
FT   DOMAIN    3759   3797       LDL-RECEPTOR CLASS A 32.
FT   DOMAIN    3798   3836       LDL-RECEPTOR CLASS A 33.
FT   DOMAIN    3842   3882       LDL-RECEPTOR CLASS A 34.
FT   DOMAIN    3883   3924       LDL-RECEPTOR CLASS A 35.
FT   DOMAIN    3928   3966       LDL-RECEPTOR CLASS A 36.
FT   DOMAIN    3968   4008       EGF-LIKE 14.
FT   DOMAIN    4009   4050       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    4156   4197       LDL-RECEPTOR CLASS B 35.
FT   DOMAIN    4199   4241       LDL-RECEPTOR CLASS B 36.
FT   DOMAIN    4244   4284       LDL-RECEPTOR CLASS B 37.
FT   DOMAIN    4332   4370       EGF-LIKE 16.
FT   DOMAIN    4379   4413       EGF-LIKE 17.
FT   SITE      4454   4460       SH3-BINDING (POTENTIAL).
FT   SITE      4457   4463       SH3-BINDING (POTENTIAL).
FT   SITE      4606   4609       SH2-BINDING (POTENTIAL).
FT   SITE      4619   4625       SH3-BINDING (POTENTIAL).
FT   SITE      4624   4630       SH3-BINDING (POTENTIAL).
FT   SITE      1743   1745       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE      4522   4527       ENDOCYTOSIS SIGNAL (POTENTIAL).
FT   SITE      4601   4606       ENDOCYTOSIS SIGNAL (POTENTIAL).
FT   DISULFID    28     40       BY SIMILARITY.
FT   DISULFID    35     53       BY SIMILARITY.
FT   DISULFID    47     62       BY SIMILARITY.
FT   DISULFID    67     80       BY SIMILARITY.
FT   DISULFID    74     93       BY SIMILARITY.
FT   DISULFID    87    103       BY SIMILARITY.
FT   DISULFID   108    120       BY SIMILARITY.
FT   DISULFID   115    133       BY SIMILARITY.
FT   DISULFID   127    142       BY SIMILARITY.
FT   DISULFID   147    157       BY SIMILARITY.

  Query Match             6.2%;  Score 94.5;  DB 1;  Length 4660;
  Best Local Similarity   23.6%;  Pred. No. 35;
  Matches   70;  Conservative   37;  Mismatches  111;  Indels   79;  Gaps   19;

Qy       1 FGKYPS---LELQPWM----------YNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTR 47
           |||       | : ||:           ||:    |||   :    | ||    :  |:
Db    4302 FGKENKEKVLVVNPWLTQVRIFHQLRYNQS---VSNPCKQVCSHLCLL----RPGGYSCA 4354

Qy      48 CMEGNPI---PDTPCQAGEEEWTTAPVPQTIMDLFQNGN-WTMQNPSPACQCSSDKIKKM 103
           | :|:         | |  |   | | |   |     || :  :|  | |:|||    :
Db    4355 CPQGSDFVTGSTVQCDAASELPVTMPPPCRCM---HGGNCYFDENELPKCKCSSGYSGE- 4410
```

```
Qy     104 LPVCPPG-AGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRY 162
           |  | : |:||               |  ::  |:   : ||  :|   : |  | |
Db    4411 --YCEVGLSRGIPP--------------GTTMA-VLLTFVIVIIVGAL---VLVGLFHY 4449

Qy     163 GGFSLGVSNTQALPPSQEVNDAIKQMKKHLKLAKDSSADRFLNSLG-RFMTGLDTRNNVK 221
                 | :| |:      : ::     ||| |      |  |  | :| |   |:
Db    4450 -------RKTGSLLPT------LPKLPSLSSLAKPSE-----NGNGVTFRSGADV--NMD 4489

Qy     222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVAL 278
           :  : |  |      :|  :   :: |: |       | :|:   |   |:|||
Db    4490 IGVSPFGPETIIDRSMAMNEHFV---MEVGKQP-----VIFENPMYAAKDNTSKVAL 4538


RESULT 14
GLN2_ORYSA
ID   GLN2_ORYSA     STANDARD;      PRT;   428 AA.
AC   P14655;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamine synthetase shoot isozyme, chloroplast precursor (EC 6.3.1.2)
DE   (Glutamate--ammonia ligase) (Clone lambda-GS31).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Kinmaze; TISSUE=Shoot;
RX   MEDLINE=91370845; PubMed=2577497;
RA   Sakamoto A., Ogawa M., Masumura T., Shibata D., Takeba G.,
RA   Tanaka K., Fujii S.;
RT   "Three cDNA sequences coding for glutamine synthetase polypeptides in
RT   Oryza sativa L.";
RL   Plant Mol. Biol. 13:611-614(1989).
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION.
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14246; CAA32462.1; -.
DR   PIR; S07471; AJRZQD.
DR   InterPro; IPR001691; GLN_synth.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family; Chloroplast; Transit peptide.
FT   TRANSIT       1     56       CHLOROPLAST (POTENTIAL).
FT   CHAIN        57    428       GLUTAMINE SYNTHETASE SHOOT ISOZYME.
SQ   SEQUENCE   428 AA;  46642 MW;  DFF1B39BFC5921FE CRC64;

  Query Match             6.1%;  Score 92.5;  DB 1;  Length 428;
  Best Local Similarity   23.6%;  Pred. No. 2.1;
  Matches   67;  Conservative   30;  Mismatches  112;  Indels   75;  Gaps   14;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
           || ||  | |:   |     || ||:  :    |: |||  |          |    | ||
Db     103 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFKDPFRGGNNILVMCDTYTPAGEPI 158

Qy      55 P----DTPCQAGEEEWTTAPVP-----QTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLP 105
             .    .
```

```
           |    :   |   :    : ||     |    | :: || :  |
Db     159 PTNKRNRAAQVFSDPKVVSQVPWFGIEQEYTLLQRDVNWPLGWP---------------- 202

Qy     106 VCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGF 165
                 || | ||     |       ||:|||   |  :                  | |
Db     203 -----VGGYPGPQGPYYCAVGSDKSFGRDISDAHYKACL----------------YAGI 240

Qy     166 SLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKV 222
           ::  :| : :|   |  | ::  :   |: ::::    :|     |  :| || :
Db     241 NISGTNGEVMPGQWEYQVGPSVGIEAGDHIWISR-YILERITEQAGVVLT-LDPKPIQGD 298

Qy     223 WFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
           | |  | |   |           ||  |||  :|:   : | ||
Db     299 W-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHDLHISAYG 341


RESULT 15
GLNC_BRANA
ID   GLNC_BRANA     STANDARD;      PRT;   428 AA.
AC   Q42624; Q9M429;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamine synthetase, chloroplast precursor (EC 6.3.1.2) (Glutamate--
DE   ammonia ligase) (GS2).
GN   GLN2 OR GLN.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=94269200; PubMed=7911583;
RA   Ochs G., Schock G., Wild A.;
RT   "Chloroplastic glutamine synthetase from Brassica napus.";
RL   Plant Physiol. 103:303-304(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Drakkar; TISSUE=Leaf;
RA   Wojtyna S., Ochs G., Wild A.;
RT   "Cloning and Sequencing of genomic fragments coding for glutamine
RT   synthetase of Brassica napus.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC       NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC       THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72751; CAA51280.1; -.
DR   EMBL; AJ271909; CAB72423.1; -.
DR   InterPro; IPR001691; GLN_synth.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family; Chloroplast; Transit peptide.
FT   TRANSIT       1     49       CHLOROPLAST (POTENTIAL).
FT   CHAIN        50    428       GLUTAMINE SYNTHETASE.
FT   CONFLICT     50     50       L -> I (IN REF. 2).
```

```
FT   CONFLICT     82     82       I -> Y (IN REF. 2).
FT   CONFLICT    263    263       G -> R (IN REF. 2).
FT   CONFLICT    338    338       S -> I (IN REF. 2).
SQ   SEQUENCE   428 AA;  47344 MW;  A0B558C64FD9B18A CRC64;

  Query Match             5.9%;  Score 90.5;  DB 1;  Length 428;
  Best Local Similarity   23.9%;  Pred. No. 3.1;
  Matches   68;  Conservative   28;  Mismatches  113;  Indels   75;  Gaps   14;

Qy       5 PSLELQPWMYNEQYTFVSNDAP-EDTGTLELLNALTKDPGFG---------TRCMEGNPI 54
           || ||  | |:   |     || ||:  :    |: :||  |         |    | ||
Db     103 PS-ELPKWNYDGSST---GQAPGEDSEVILYPQAIFRDPFRGGNNILVICDTYTPAGEPI 158

Qy      55 P-DTPCQAGE--------EEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLP 105
           | :   :| |        ||     : |    |  | || :  |
Db     159 PTNKRARAAEIFSNKKVNEEIPWFGIEQEYTLLQPNVNWPLGWP--------------- 202

Qy     106 VCPPGAGGLPPPQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGF 165
             |  | ||         :   ||:|||   |  :                 | |
Db     203 -----VGAYPGPQGPYYCGVGAEKSWGRDISDAHYKACL----------------YAGI 240

Qy     166 SLGVSNTQALPPSQE--VNDAIK-QMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKV 222
           ::  :| : :|   |  |  ::  :   |:  |:    :|     |  :| || :
Db     241 NISGTNGEVMPGQWEFQVGPSVGIEAGDHVWCAR-YLLERITEQAGVVLT-LDPKPIEGD 298

Qy     223 WFNNKGWHAISSF--------LNVINNAILRANLQKGENPSHYG 258
           | |  | |   |           ||  |||  :|:  |: | ||
Db     299 W-NGAGCHTNYSTKSMREDGGFEVIKKAILNLSLRHMEHISAYG 341


Search completed: February  4, 2003, 09:39:35
Job time : 19 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2003, 09:38:24 ; Search time 33 Seconds
(without alignments)
1773.254 Million cell updates/sec

Title: US-09-704-272-6
Perfect score: 1525
Sequence: 1 FGKYPSLELQPWMYNEQYTF.....PLNLTKQQLSEVALMTTSVD 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1235.5	81.0	2260	13	Q8UVV4	Q8uvv4 gallus gall
2	731.5	48.0	2281	6	O02698	O02698 bos taurus
3	724.5	47.5	2310	11	O35600	O35600 mus musculu
4	664	43.5	2159	11	Q91V24	Q91v24 mus musculu
5	662.5	43.4	2146	4	Q9BZC4	Q9bzc4 homo sapien
6	658.5	43.2	2008	4	Q96S58	Q96s58 homo sapien
7	658.5	43.2	2146	4	Q9NR73	Q9nr73 homo sapien
8	267	17.5	1529	4	Q9UPU0	Q9upu0 homo sapien
9	267	17.5	2436	4	Q9HC28	Q9hc28 homo sapien
10	262	17.2	867	4	Q96HC2	Q96hc2 homo sapien
11	259.5	17.0	2434	11	Q9ESR9	Q9esr9 rattus norv
12	250	16.4	2277	4	Q96JT3	Q96jt3 homo sapien
13	215.5	14.1	1547	5	O01790	O01790 caenorhabdi
14	114.5	7.5	961	5	Q8SWX4	Q8swx4 drosophila
15	114.5	7.5	1878	5	Q9VD86	Q9vd86 drosophila
16	114	7.5	1500	5	Q9VRF4	Q9vrf4 drosophila
17	102	6.7	904	12	Q9IWU4	Q9iwu4 human herpe
18	100	6.6	345	17	Q8TUM6	Q8tum6 methanosarc
19	98	6.4	877	16	Q99R07	Q99r07 staphylococ
20	98	6.4	913	2	O86476	O86476 staphylococ
21	97.5	6.4	432	10	Q9SPJ1	Q9spj1 juglans nig
22	97.5	6.4	828	5	Q9W292	Q9w292 drosophila
23	97	6.4	1452	12	Q8UYV8	Q8uyv8 strawberry
24	96	6.3	803	5	Q8T2J0	Q8t2j0 dictyosteli
25	95	6.2	1256	5	Q95YA9	Q95ya9 caenorhabdi
26	91.5	6.0	430	10	Q9SEX6	Q9sex6 canavalia l
27	91	6.0	274	10	Q9LWA7	Q9lwa7 oryza sativ
28	91	6.0	302	5	Q8SU46	Q8su46 encephalito
29	91	6.0	459	16	Q9KDR5	Q9kdr5 bacillus ha
30	90.5	5.9	431	10	Q9AWA8	Q9awa8 beta vulgar
31	90.5	5.9	433	10	Q9XFJ1	Q9xfj1 mesembryant
32	90	5.9	649	11	Q9D296	Q9d296 mus musculu
33	90	5.9	773	16	Q98R04	Q98r04 mycoplasma
34	89.5	5.9	432	8	Q95AG1	Q95ag1 glycine max
35	89.5	5.9	649	10	Q9CAL3	Q9cal3 arabidopsis
36	89.5	5.9	1516	4	Q8TEL8	Q8tel8 homo sapien
37	89.5	5.9	1534	4	O75093	O75093 homo sapien
38	89	5.8	449	16	Q8RG05	Q8rg05 fusobacteri
39	89	5.8	903	12	Q69076	Q69076 human herpe
40	88.5	5.8	865	16	Q8YW75	Q8yw75 anabaena sp
41	88.5	5.8	1071	5	Q17405	Q17405 caenorhabdi
42	88.5	5.8	1650	11	Q9QVT6	Q9qvt6 rattus sp.
43	88	5.8	365	5	O97273	O97273 plasmodium
44	88	5.8	748	17	Q9HLG7	Q9hlg7 thermoplasm
45	88	5.8	791	3	Q08817	Q08817 saccharomyc

ALIGNMENTS

RESULT 1
Q8UVV4
ID Q8UVV4 PRELIMINARY; PRT; 2260 AA.
AC Q8UVV4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATP-binding cassette transporter 1.
GN ABCA1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Attie A.D., Brooks-Wilson A.R., Walker D., McManus B.,
RA Gray-Kellar M.P., MacDonald M.L.E., Roomp K., Tebon A., Zhang L.-H.,
RA Mulligan J., Sensen C., Bitgood J.J., Cook M.E., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Cholesterol Ester Accumulation in Hepatocytes and Intestinal Lamina
RT Propria Caused by an ABCA1 Mutation in WHAM Chickens.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF362377; AAL56247.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transportr; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 2260 AA; 254070 MW; 19D137F342F98662 CRC64;

Query Match 81.0%; Score 1235.5; DB 13; Length 2260;
Best Local Similarity 78.9%; Pred. No. 7.6e-97;
Matches 224; Conservative 33; Mismatches 26; Indels 1; Gaps 1;

Qy 1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60

```
Db   1371 FGKYPSLELQPWMYDEQYTFISNDAPEDAGTQKLLDALLNKPGFGTRCMQGHSIPDTPCT 1430

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
          |::|||||| || ::::: : |||:|:||||:|:||::||||||||||||||||||||:
Db   1431 VGQKEWTTASVPDSVLEILR-GNWSMENPSPSCECSNEKIKKMLPVCPPGAGGLPPPQRE 1489

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          |:||||||:|||||||||||||| ||| ||||||||||||||||||  ::  |||| |
Db   1490 QDTADILQNLTGRNISDYLVKTYAQIIGKSLKNKIWVNEFRYGGFSLGARSSHVLPPSNE 1549

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
          | |||||:|| |:||: || ||||||:|  || ||||:|||||||||||||||:|||||||
Db   1550 VTDAIKQVKKILELAQGSSGDRFLNNLASFMKGLDTKNNVKVWFNNKGWHAIASFLNVIN 1609

Qy    241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
          |||||||||:|:||| ||||||||||||||||||||||||||||
Db   1610 NAILRANLQQGKNPSAYGITAFNHPLNLTKQQLSEVALMTTSVD 1653


RESULT 2
O02698
ID   O02698      PRELIMINARY;      PRT;  2281 AA.
AC   O02698;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ABC transporter.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINAL ROD CELL;
RX   MEDLINE=97248596; PubMed=9092582;
RA   Illing M., Molday L.L., Molday R.S.;
RT   "The 220-kDa rim protein of retinal rod outer segments is a member of
RT   the ABC transporter superfamily.";
RL   J. Biol. Chem. 272:10303-10310(1997).
DR   EMBL; U90126; AAC48716.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01257; rim_protein; 1.
KW   ATP-binding.
SQ   SEQUENCE   2281 AA;  257228 MW;  71CD404C98F7A079 CRC64;

  Query Match             48.0%;  Score 731.5;  DB 6;  Length 2281;
  Best Local Similarity   47.2%;  Pred. No. 1.4e-53;
  Matches  143;  Conservative   38;  Mismatches   83;  Indels   39;  Gaps    4;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||:||:| | |||| :|||| | | |:      | : |   |||| ||::   :|: ||
Db   1395 FGEYPALTLHPWMYGQQYTFFSMDQPDSEWLSALADVLVNKPGFGNRCLKEEWLPEFPC- 1453

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
            | |  |   :  | |   ||   |||:|:||: :   ||| || ||||||||||
Db   1454 GNSSPWKTPSVSPDVTHLLQQQKWTADQPSPSCRCSTREKLTMLPECPEGAGGLPPPQRI 1513

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | : :|||||| ||:||:||||| :|  |||:| |||| |||| |:|      ||
Db   1514 QRSTEILQDLTDRNVSDFLVKTYPALIRSSLKSKFWVNEQRYGGISVG----GKLPAPPF 1569

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGR----------FMTGLDTRNNVK 221
            :|:                ||: ||:                  |:  |:| :|:|
Db   1570 TGEALV--------------GFLSDLGQLMNVSGGPMTREAAKEMPAFLKQLETEDNIK 1614

Qy    222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTT 281
          |||||||||: ||||| :|||||:| | :||  |||| : ||||||:||||: ::||
Db   1615 VWFNNKGWHALVSFLNVAHNAILRASLHKDKNPEEYGITVISQPLNLTKEQLSEITVLTT 1674

Qy    282 SVD 284
          |||
Db   1675 SVD 1677


RESULT 3
O35600
ID   O35600      PRELIMINARY;      PRT;  2310 AA.
AC   O35600;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP-binding cassette transporter.
GN   ABCA4 OR ABCR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=97345663; PubMed=9202155;
RA   Azarian S.M., Travis G.H.;
RT   "The photoreceptor rim protein is an ABC transporter encoded by the
RT   gene for recessive Stargardt's disease (ABCR).";
RL   FEBS Lett. 409:247-252(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Azarian S.M., Travis G.H.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF000149; AAC23916.1; -.
DR   MGD; MGI:109424; Abca4.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR01257; rim_protein; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2310 AA;  260207 MW;  8370C6C8A62EF294 CRC64;

  Query Match             47.5%;  Score 724.5;  DB 11;  Length 2310;
  Best Local Similarity   47.2%;  Pred. No. 5.8e-53;
  Matches  143;  Conservative   39;  Mismatches   82;  Indels   39;  Gaps    4;

Qy      1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
          ||::|:| | |||| |||| | | :      | : |   |||| ||::   :|: ||
Db   1396 FGEFPALTLHPWMYGHQYTFFSMDEPNNEHLEVLADVLLNRPGFGNRCLKEEWLPEYPC- 1454

Qy     61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
            | |  |   | |||   ||  :|||:|:||: :   ||| || ||||||||||
Db   1455 INATSWKTPSVSPNITHLFQKQKWTAAHPSPSCKCSTREKLTMLPECPEGAGGLPPPQRT 1514

Qy    121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
          | : ::||||| ||||||||||  :|  |||:| |||| |||| |:|      |:| | |
Db   1515 QRSTEVLQDLTNRNISDYLVKTYPALIRSSLKSKFWVNEQRYGGISIG-GKLPAIPISGE 1573

Qy    181 VNDAIKQMKKHLKLAKDSSADRFLNSLGR------------------FMTGLDTRNNVK 221
                        :  ||: ||:                    |:  |:| :|:|
Db   1574 ------------------ALVGFLSGLGQMMNVSGGPVTREASKEMLDFLKHLETTDNIK 1615

Qy    222 VWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTT 281
          |||||||||: ||||| :|||||:| :  :|  |||| : ||||||:|||:: ::||
Db   1616 VWFNNKGWHALVSFLNVAHNAILRASLPRDRDPEEYGITVISQPLNLTKEQLSDITVLTT 1675

Qy    282 SVD 284
          |||
Db   1676 SVD 1678
```

```
RESULT 4
Q91V24
ID   Q91V24      PRELIMINARY;      PRT;  2159 AA.
AC   Q91V24;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP-binding cassette transporter sub-family A member 7.
GN   ABCA7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OLl29, AND DBA/2;
RX   MEDLINE=21328888; PubMed=11435699;
RA   Broccardo C., Osorio J., Luciani M.-F., Schriml L.M., Prades C.,
RA   Shulenin S., Arnould I., Naudin L., Lafargue C., Rosier M., Jordan B.,
RA   Mattei M.G., Dean M., Denefle P., Chimini G.;
RT   "Comparative analysis of the promoter structure and genomic
RT   organization of the human and mouse ABCA7 gene encoding a novel ABCA
RT   transporter.";
RL   Cytogenet. Cell Genet. 92:264-270(2001).
DR   EMBL; AF287142; AAK56863.1; -.
DR   EMBL; AF287141; AAK56862.1; -.
DR   MGD; MGI:1351646; Abca7.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR002016; Peroxidase.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2159 AA;  236882 MW;  CD2BE3FE0D8B822B CRC64;

  Query Match             43.5%;  Score 664;  DB 11;  Length 2159;
  Best Local Similarity   45.3%;  Pred. No. 8.3e-48;
  Matches  129;  Conservative   47;  Mismatches  107;  Indels    2;  Gaps    2;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           . ||:|| |:| | ||  | :| | ||| |   ::|| ||  : |     |:       : |
Db    1266 FGQYPPLQLSPAMYGPQVSFFSEDAPGDPNRMKLLEALLGEAGLQEPSMQDKDARGSECT 1325

Qy      61 AGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQR 119
             : | | ||  :  :  :|||| ::|||||||||    :::|| || |||| ||||
Db    1326 HSLACYFTVPEVPPDVASILASGNWTPESPSPACQCSQPGARRLLPDCPAGAGGPPPPQA 1385

Qy     120 KQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQ 179
             :::|:|||||:||:|||||  :: : || | ||:| ||||||||      ||
Db    1386 VAGLGEVVQNLTGRNVSDFLVKTYPSLVRRGLKTKKWVDEVRYGGFSLG-GRDPDLPTGH 1444

Qy     180 EVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVI 239
           ||   : :::  |     :: || ||:| ::  ||| ||::|:|||||||||: :|:|
Db    1445 EVVRTLAEIRALLSPQPGNALDRILNNLTQWALGLDARNSLKIWFNNKGWHAMVAFVNRA 1504

Qy     240 NNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           || :| | |  |     : ||  |||||||:|||  |: :|||
Db    1505 NNGLLHALLPSGPVRHAHSITTLNHPLNLTKEQLSEATLIASSVD 1549


RESULT 5
Q9BZC4
ID   Q9BZC4      PRELIMINARY;      PRT;  2146 AA.
AC   Q9BZC4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ABC transporter member 7.
GN   ABCA7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21328888; PubMed=11435699;
RA   Broccardo C., Osorio J., Luciani M.-F., Schriml L.M., Prades C.,
RA   Shulenin S., Arnould I., Naudin L., Lafargue C., Rosier M., Jordan B.,
RA   Mattei M.G., Dean M., Denefle P., Chimini G.;
RT   "Comparative analysis of the promoter structure and genomic
RT   organization of the human and mouse ABCA7 gene encoding a novel ABCA
RT   transporter.";
RL   Cytogenet. Cell Genet. 92:264-270(2001).
DR   EMBL; AF328787; AAK00959.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2146 AA;  234306 MW;  2391728D5AD97E75 CRC64;

  Query Match             43.4%;  Score 662.5;  DB 4;  Length 2146;
  Best Local Similarity   44.7%;  Pred. No. 1.1e-47;
  Matches  127;  Conservative   50;  Mismatches   96;  Indels   11;  Gaps    2;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db    1263 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1312

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
                ::   ||  :  :  :|||| ::||||||||    :::|| ||  ||| ||||
Db    1313 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSRPGARRLLPDCPAAAGGPPPPQAV 1372

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
             :.:::|:|||||:||:|||||| ::: : || | |||| |||||||      ||  ||
Db    1373 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1431

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           :  ::::   |      : || | :|  :   || ::::|:|||||||:: :|:|  :
Db    1432 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1491

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           ||||||:|  |     : ||  |||||||:|||| ||| :|||
Db    1492 NAILRAHLPPGPARHAHSITTLNHPLNLTKEQLSEAALMASSVD 1535


RESULT 6
Q96S58
ID   Q96S58      PRELIMINARY;      PRT;  2008 AA.
AC   Q96S58;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ABCA-SSN.
GN   ABCA7/ABCA-SSN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21255283; PubMed=11355874;
RA   Tanaka A., Ikeda Y., Abe-Dohmae S., Arakawa R., Sadanami K.,
RA   Kidera A., Nakagawa S., Nagase T., Aoki R., Kioka N., Amachi T.,
RA   Yokoyama S., Ueda K.;
RT   "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain
RT   Homologous to an Epitope of Sjogren's Syndrome.";
RL   Biochem. Biophys. Res. Commun. 283:1019-1025(2001).
DR   EMBL; AB055390; BAB62294.1; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001899; Gram_pos_anchor.
```

```
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ   SEQUENCE   2008 AA;  218617 MW;  226FF85C24230B90 CRC64;

  Query Match             43.2%;  Score 658.5;  DB 4;  Length 2008;
  Best Local Similarity   44.4%;  Pred. No. 2.2e-47;
  Matches  126;  Conservative   50;  Mismatches   97;  Indels   11;  Gaps    2;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db    1125 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1174

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  :  :|||| ::|||||||||    :::|| ||  ||| ||||
Db    1175 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSQPGARRLLPDCPAAAGGPPPPQAV 1234

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db    1235 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1293

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           :  ::::   |      : || | :|  :   || ::::|:||||||||:: :|:|  :
Db    1294 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1353

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           ||||||:|  |      : ||  |||||||||:|| | ||| :|||
Db    1354 NAILRAHLPPGRARHAHSITTLNHPLNLTKEQLFEAALMASSVD 1397


RESULT 7
Q9NR73
ID   Q9NR73       PRELIMINARY;      PRT;  2146 AA.
AC   Q9NR73;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Macrophage ABC transporter.
GN   ABCA7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20334305; PubMed=10873640;
RA   Kaminski W.E., Orso E., Diederich W., Klucken J., Drobnik W.,
RA   Schmitz G.;
RT   "Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette
RT   Transporter (ABCA7).";
RL   Biochem. Biophys. Res. Commun. 273:532-538(2000).
DR   EMBL; AF250238; AAF85794.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2146 AA;  234469 MW;  679B16EB2D75FF0D CRC64;

  Query Match             43.2%;  Score 658.5;  DB 4;  Length 2146;
  Best Local Similarity   44.4%;  Pred. No. 2.4e-47;
  Matches  126;  Conservative   50;  Mismatches   97;  Indels   11;  Gaps    2;

Qy       1 FGKYPSLELQPWMYNEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPCQ 60
           || ||:| | | ||  | :| | ||| | |   || || :: |          : : | |
Db    1263 FGHYPALRLSPTMYGAQVSFFSEDAPGDPGRARLLEALLQEAG----------LEEPPVQ 1312

Qy      61 AGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRK 120
             ::   ||  :  :  :|||| ::|||||||||    :::|| ||  ||| ||||
Db    1313 HSSHRFSAPEVPAEVAKVLASGNWTPESPSPACQCSQPGARRLLPDCPAAAGGPPPPQAV 1372

Qy     121 QNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQE 180
           : :::|:|||||:||:||||| ::: : || | |||| ||||||||      ||  ||
Db    1373 TGSGEVVQNLTGRNLSDFLVKTYPRLVRQGLKTKKWVNEVRYGGFSLG-GRDPGLPSGQE 1431

Qy     181 VNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVIN 240
           :  ::::   |      : || | :|  :   || ::::|:||||||||:: :|:|  :
Db    1432 LGRSVEELWALLSPLPGGALDRVLKNLTAWAHSLDAQDSLKIWFNNKGWHSMVAFVNRAS 1491

Qy     241 NAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           ||||||:|  |      : ||  |||||||||:|| | ||| :|||
Db    1492 NAILRAHLPPGRARHAHSITTLNHPLNLTKEQLFEAALMASSVD 1535


RESULT 8
Q9UPU0
ID   Q9UPU0       PRELIMINARY;      PRT;  1529 AA.
AC   Q9UPU0; Q9NSL2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   KIAA1062 protein (Fragment).
GN   KIAA1062 OR DKFZP547P193.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
RN   [2]
RP   SEQUENCE OF 1157-1529 FROM N.A.
RC   TISSUE=BRAIN;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB028985; BAA83014.1; -.
DR   EMBL; AL162060; CAB82398.1; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   1529 AA;  170059 MW;  BCED65F5E4AACA1A CRC64;

  Query Match             17.5%;  Score 267;  DB 4;  Length 1529;
  Best Local Similarity   29.2%;  Pred. No. 6.3e-14;
  Matches   80;  Conservative   35;  Mismatches   73;  Indels   86;  Gaps   12;

Qy      44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
           | : |:|                :| | |:|                || | ||:|| :
Db     665 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 724

Qy      72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
           |: : :  :            | ||:         ||   ||  ||| :  | ||| |:|
Db     725 PRLVREPVR-----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDIT 768

Qy     132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
           | |:|:||: |  :               |||  : |  |     |:     |   ::|
Db     769 GHNVSEYLLFTSDRF-----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 814
```

```
Qy    192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                              :  |   :|::||||:|:: ::|| :|||||||| |
Db    815 ----------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 852

Qy    252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
          : ||: ||||  |||:| |   || :  |  | |
Db    853 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 886


RESULT 9
Q9HC28
ID   Q9HC28      PRELIMINARY;      PRT;  2436 AA.
AC   Q9HC28;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP-binding cassette sub-family A member 2 (ABC transporter
DE   ABCA2).
GN   ABCA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Vulevic B., Chen Z., Davis W. Jr., Walsh E.S., Tew K.D.;
RT   "Cloning and characterization of human ABCA2.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   PubMed=11178988;
RA   Kaminski W.E., Piehler A., Pullmann K., Porsch-Ozcurumez M., Duong C.,
RA   Bared G.M., Buchler C., Schmitz G.;
RT   "Complete Coding Sequence, Promoter Region, and Genomic Structure of
RT   the Human ABCA2 Gene and Evidence for Sterol-Dependent Regulation in
RT   Macrophages.";
RL   Biochem. Biophys. Res. Commun. 281:249-258(2001).
DR   EMBL; AF178941; AAG09372.1; -.
DR   EMBL; AF327657; AAK14334.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000572; Euk_Mb_oxred.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR   PROSITE; PS00559; MOLYBDOPTERIN_EUK; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2436 AA;  269955 MW;  E044A3AF14EA25D1 CRC64;

  Query Match             17.5%;  Score 267;  DB 4;  Length 2436;
  Best Local Similarity   29.2%;  Pred. No. 1.2e-13;
  Matches   80;  Conservative   35;  Mismatches   73;  Indels   86;  Gaps   12;

Qy     44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
          | : |:|                :| | |:|              || | ||:|| :
Db   1572 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 1631

Qy     72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
          |: : :  :          | ||:         ||   ||  ||| :  | ||| |:|
Db   1632 PRLVREPVR----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVTGDILTDIT 1675

Qy    132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
          | |:|:||: |  :              |||  : |  |     |:     |   ::|
Db   1676 GHNVSEYLLFTSDRF----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 1721

Qy    192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                              :  |   :|::||||:|:: ::|| :|||||||| |
Db   1722 ----------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 1759

Qy    252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
          : ||: ||||  |||:| |   || :  |  | |
Db   1760 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 1793


RESULT 10
Q96HC2
ID   Q96HC2      PRELIMINARY;      PRT;  867 AA.
AC   Q96HC2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Similar to KIAA1062 protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008755; AAH08755.1; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transportr; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER       1       1
SQ   SEQUENCE   867 AA;  96734 MW;  DCF6B6A90074C085 CRC64;

  Query Match             17.2%;  Score 262;  DB 4;  Length 867;
  Best Local Similarity   28.8%;  Pred. No. 8e-14;
  Matches   79;  Conservative   35;  Mismatches   74;  Indels   86;  Gaps   12;

Qy     44 FGTRCMEG--------------NPIP-DTPCQ--------------AGEEEWTTAP-V 71
          | : |:|                :| | |:|              || | ||:|| :
Db      3 FDSMCLESFTQGLPLSNFVPPPPSPAPSDSPASPDEDLQAWNVSLPPTAGPEMWTSAPSL 62

Qy     72 PQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLT 131
          |: : :  :          | ||:         ||   ||  ||| :  | ||| |:|
Db     63 PRLVREPVR----------CTCSAQGTGFS---CPSSVGG-HPPQMRVVAGDILTDIT 106

Qy    132 GRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVNDAIKQMKKH 191
          | |:|:||: |  :              |||  : |  |     |:     |   ::|
Db    107 GHNVSEYLLFTSDRF----------RLHRYGAITFG--NVLKSIPASFGTRAPPMVRK- 152

Qy    192 LKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKG 251
                              :  |   :|::||||:|:: ::|| :|||||||| |
Db    153 ----------------------IAVRRAAQVFYNNKGYHSMPTYLNSLNNAILRANLPKS 190

Qy    252 E-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
          : ||: ||||  |||:| |   || :  |  | |
Db    191 KGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 224


RESULT 11
Q9ESR9
ID   Q9ESR9      PRELIMINARY;      PRT;  2434 AA.
AC   Q9ESR9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ABC2.
GN   ABC2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20427713; PubMed=10970803;
RA   Zhao L., Zhou C., Tanaka A., Nakata M., Hirabayashi T., Amachi T.,
RA   Shioda S., Ueda K., Inagaki N.;
RT   "Cloning, characterization and tissue distribution of the rat ATP-
RT   binding cassette (ABC) transporter ABC2/ABCA2.";
RL   Biochem. J. 350:865-872(2000).
DR   EMBL; AB037937; BAB16596.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2434 AA;  270925 MW;  CD424A9C4F63513F CRC64;

  Query Match             17.0%;  Score 259.5;  DB 11;  Length 2434;
  Best Local Similarity   32.3%;  Pred. No. 5.2e-13;
  Matches   74;  Conservative   29;  Mismatches   71;  Indels   55;  Gaps    9;

Qy      58 PCQAGEEEWTTAP-VPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPP 116
           |  || | || || :|: : :  :             | ||:          ||   ||  |
Db    1617 PPTAGPETWTWAPSLPRLVHEPVR-----------CTCSAQGTGFS---CPSSVGG-HP 1660

Qy     117 PQRKQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALP 176
           || :  | ||| |:|| |:|:||: |  :                |||  : |  | |
Db    1661 PQMRVVTGDILTDITGHNVSEYLLFTSDRF-----------RLHRYGAITFG--NIQKSI 1707

Qy     177 PSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLDTRNNVKVWFNNKGWHAISSFL 236
           |:       ::|                    :  |   :| :||||:|:: ::|
Db    1708 PAPIGTRTPLMVRK---------------------IAVRRVAQVLYNNKGYHSMPTYL 1744

Qy     237 NVINNAILRANLQKGE-NPSHYGITAFNHPLNLTKQQLS-EVALMTTSV 283
           | :||||||||| | : ||: |||| |||:| |   || :  |  | |
Db    1745 NSLNNAILRANLPKSKGNPAAYGITVTNHPMNKTSASLSLDYLLQGTDV 1793


RESULT 12
Q96JT3
ID   Q96JT3      PRELIMINARY;      PRT;  2277 AA.
AC   Q96JT3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP-binding cassette transporter family A member 12.
GN   ABCA12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RA   Bonner T.I., Moses T., Detera-Wadleigh S.;
RT   "A retinal cDNA for the ATP-binding cassette transporter ABCA12.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033486; AAK54355.1; -.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR003583; HHH_1.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00278; HhH1; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW   ATP-binding.
SQ   SEQUENCE   2277 AA;  256970 MW;  EDA2F00280361E2D CRC64;

  Query Match             16.4%;  Score 250;  DB 4;  Length 2277;
  Best Local Similarity   26.0%;  Pred. No. 3.1e-12;
  Matches   75;  Conservative   49;  Mismatches   81;  Indels   84;  Gaps   14;

Qy       4 YPSLELQPWMY--NEQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPDTPC-- 59
           || ::: | :|  :|| | :|  |    |  |::|:  ||    |: :  | |
Db    1456 YPEIQISPSLYGTSEQTAFYANYHP---STEALVSAMWDFPGIDNMCLNTS---DLQCLN 1509

Qy      60 QAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQR 119
           :  |:| |:  | |  :              | || :    :  ||       || |
Db    1510 KDSLEKWNTSGEPITNFGV-------------CSCSEN-----VQECP--KFNYSPPHR 1548

Qy     120 KQNTADILQDLTGRNISDYLVKTYVQIIAKSLKNKIWVNEF---RYGGFSLGVSNTQALP 176
           :  :: :: :|||: : :||: |              |||   ||||:| |
Db    1549 RTYSSQVIYNLTGQRVENYLIST-------------ANEFVQKRYGGWSFG-------- 1586

Qy     177 PSQEVNDAIKQMKKHLKLAKDSSADRFLNSLGRFMTGLD-TRNNVKVWFNNKGWHAISSF 235
                       | | ||   |         :||:   |   |||:: :|:|:: ::
Db    1587 --------------LPLTKDLRFD---------ITGVPANRTLAKVWYDPEGYHSLPAY 1622

Qy     236 LNVINNAILRANLQKGENPSHYGITAFNHPLNLTKQQLSEVALMTTSVD 284
           || :|| :|| |: | :   | ||  ::||    : |  | | ::: :|
Db    1623 LNSLNNFLLRVNMSKYDAARH-GIIMYSHPYPGVQDQ--EQATISSLID 1668


RESULT 13
O01790
ID   O01790      PRELIMINARY;      PRT;  1547 AA.
AC   O01790;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 173.7 kDa protein.
GN   F12B6.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Pauley A., Maggi L.;
RT   "The sequence of C. elegans cosmid F12B6.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF003138; AAK21369.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1547 AA;  173682 MW;  A532D93977006C67 CRC64;

  Query Match             14.1%;  Score 215.5;  DB 5;  Length 1547;
  Best Local Similarity   24.2%;  Pred. No. 1.7e-09;
  Matches   86;  Conservative   44;  Mismatches  105;  Indels  121;  Gaps   15;

Qy       5 PSLELQPWMYNEQYTFVSN--DAPEDTGTLELLNALTKDPGFGTRCMEG--NPIPDT--- 57
           | | |:  :      |  |  |  |::   ::|:|:   || | || :   | : ||
Db     605 PPLPLETSIMGNHSDFYVNSWDTAENSTANDILHAMFSSPGTGPRCAKDVPNDLLDTMRR 664

Qy      58 ----------------------------PCQ--AGEEEWT--------TAPVPQTIMDL- 78
                                        ||   || ::|         ||:     |
Db     665 ELMFRNRYGFGRNKPAPGVDKDSVDNEYQCQNIQGEFDYTEDISNATYNAPIYCGCEDFG 724

Qy      79 ----FQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPQRKQNTADILQDLTGRN 134
```

```
              ::  |   |                                  : || | : ||||||
Db    725 WNCTLEDWKWNETN----------------------------WLRLNTTDRIFDLTGRN 755

Qy    135 ISDY-LVKTYVQIIAKSLKNKIWVNEFRYGGFSLGVSNTQALPPSQEVND--------AI 185
          :: : |:  : |:   :          |  ||||||  | :|   ||   |         |
Db    756 LTQFRLITRFAQLANTTA-------PFFLGGFSLGHVNQRA--QSQADIDTSKRGWLETI 806

Qy    186 KQMKKHLKL-------------------AKDSSADRFLNSLGRFMTGLDTRNNVKVWFNN 226
          | : : :::                   |:: : :: :| |   :  || | ||||||||
Db    807 KDIAQSMRIINLNTTGIEPATPKVLDPFAQNITLNQVVNDL---LQNLDVRENVKVWFNN 863

Qy    227 KGWHAISSFLNVINNAILRANLQKGENPSHYGITAFNHPLNLT-KQQLSEVALMTT 281
          | |       |:::||:||       :|   ||   |||:| |  | | : ||  |
Db    864 KIWPGFPIASNILSNALLRQE-DYAIDPEDLGILTMNHPMNKTISQTLDQNALKFT 918


RESULT 14
Q8SWX4
ID   Q8SWX4      PRELIMINARY;      PRT;   961 AA.
AC   Q8SWX4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   GH24371p.
GN   CG5839.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY095013; AAM11341.1; -.
SQ   SEQUENCE   961 AA;  108360 MW;  0337B20128DA5EDE CRC64;

  Query Match             7.5%;  Score 114.5;  DB 5;  Length 961;
  Best Local Similarity   21.0%;  Pred. No. 0.42;
  Matches   62;  Conservative   45;  Mismatches   81;  Indels  107;  Gaps   14;

Qy      6 SLELQPWMYN---------EQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPD 56
          :|: ||| ::          : | ||| |    :  ||||  || :  :
Db    616 ALKSQPWRFHPRNRAQLLYDTYIFVSTDRLSHSILLELLGYLTNEQQYA----------- 664

Qy     57 TPCQAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPP 116
                  |:||    |: | :  |: :  |          | :|:::
Db    665 --------PWSTANTILTVYDRYLRGDDSYYN-------FQDFVKRLI------------ 697

Qy    117 PQRKQNTADILQDLTGRN--ISDYLVKTYVQIIAKSL-------------KNKIWVNEFR 161
                  | :  |  | |    :: :  |::|:  ||               ||:  :|:
Db    698 --------DPIFDKIGVNEIPGEHYLNNYLRIVLVSLACQVGSDDCYNQSANKL--SEYL 747

Qy    162 YGGFSLGVS-NTQALPPS--QEVNDAIKQMKKHLKLAKDSSADR--FLNSLG-------- 208
          | | ::  :  |||        |:   ::: | |:   | ||  |::|||
Db    748 YNGTAIEATLKTQAYCAGLRSTTNEIYSRVQSDL-LSSSDSTDRSLFISSLGCSGSTSQL 806

Qy    209 --RFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITA 261
                |||  |::          : |   :::|:|  |          |  |:||
Db    807 LDFLRLSLDTNNSL----------SYSERTSLLNSAYSR---------SEIGLTA 842


RESULT 15
Q9VD86
ID   Q9VD86      PRELIMINARY;      PRT;  1878 AA.
AC   Q9VD86;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CG5839 protein.
GN   CG5839.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003736; AAF55912.1; -.
DR   MEROPS; M01.013; -.
DR   FlyBase; FBgn0038898; CG5839.
DR   InterPro; IPR001930; Ala_peptase.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01433; Peptidase_M1; 2.
DR   PRINTS; PR00756; ALADIPTASE.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1878 AA;  211470 MW;  2E2E4FF706373FE7 CRC64;

  Query Match             7.5%;  Score 114.5;  DB 5;  Length 1878;
  Best Local Similarity   21.0%;  Pred. No. 1;
  Matches   62;  Conservative   45;  Mismatches   81;  Indels  107;  Gaps   14;

Qy      6 SLELQPWMYN---------EQYTFVSNDAPEDTGTLELLNALTKDPGFGTRCMEGNPIPD 56
          :|: ||| ::          : | ||| |    :  ||||  || :  :
Db    616 ALKSQPWRFHPRNRAQLLYDTYIFVSTDRLSHSILLELLGYLTNEQQYA----------- 664

Qy     57 TPCQAGEEEWTTAPVPQTIMDLFQNGNWTMQNPSPACQCSSDKIKKMLPVCPPGAGGLPP 116
                  |:||    |: | :  |: :  |          | :|:::
Db    665 --------PWSTANTILTVYDRYLRGDDSYYN-------FQDFVKRLI------------ 697
```

```
Qy    117 PQRKQNTADILQDLTGRN--ISDYLVKTYVQIIAKSL------------KNKIWVNEFR 161
                  | : |  | |    :: :  |::|:  ||              ||:  :|:
Db    698 --------DPIFDKIGVNEIPGEHYLNNYLRIVLVSLACQVGSDDCYNQSANKL--SEYL 747

Qy    162 YGGFSLGVS-NTQALPPS--QEVNDAIKQMKKHLKLAKDSSADR--FLNSLG-------- 208
          | | ::  :  |||         |:   :::  | |:   | ||  |::|||
Db    748 YNGTAIEATLKTQAYCAGLRSTTNEIYSRVQSDL-LSSSDSTDRSLFISSLGCSGSTSQL 806

Qy    209 --RFMTGLDTRNNVKVWFNNKGWHAISSFLNVINNAILRANLQKGENPSHYGITA 261
                |||  |::          : |   :::|:|  |          |  |:||
Db    807 LDFLRLSLDTNNSL----------SYSERTSLLNSAYSR---------SEIGLTA 842
```

Search completed: February 4, 2003, 09:40:17
Job time : 43 secs